Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir

Heather L Wander[1], R. Quinn Thomas[1], Tadhg N Moore[1], Mary E. Lofton[1], Adrienne Breef-Pilz[1], and Cayelan Carey[1]

[1]Affiliation not available

May 25, 2023

Abstract

Ecosystems around the globe are experiencing increased variability due to land use and climate change. In response, ecologists are increasingly using near-term, iterative ecological forecasts to predict how ecosystems will change in the future. To date, many near-term, iterative forecasting systems have been developed using high temporal frequency (minute to hourly resolution) data streams for assimilation. However, this approach may be cost-prohibitive or impossible for forecasting ecological variables that lack high-frequency sensors or have high data latency (i.e., a delay before data are available for modeling after collection). To explore the effects of data assimilation frequency on forecast skill, we developed water temperature forecasts for a eutrophic drinking water reservoir and conducted data assimilation experiments by selectively withholding observations to examine the effect of data availability on forecast accuracy. We used in-situ sensors, manually collected data, and a calibrated water quality ecosystem model driven by forecasted weather data to generate future water temperature forecasts using FLARE (Forecasting Lake And Reservoir Ecosystems), an open-source water quality forecasting system. We tested the effect of daily, weekly, fortnightly, and monthly data assimilation on the skill of 1 to 35-day-ahead water temperature forecasts. We found that forecast skill varied depending on the season, forecast horizon, depth, and data assimilation frequency, but overall forecast performance was high, with a mean 1-day-ahead forecast root mean square error (RMSE) of 0.94°C, mean 7-day RMSE of 1.33°C, and mean 35-day RMSE of 2.15°C. Aggregated across the year, daily data assimilation yielded the most skillful forecasts at 1-7-day-ahead horizons, weekly data assimilation resulted in the most skillful forecasts at 8-35-day-ahead horizons. Within a year, daily to fortnightly data assimilation substantially outperformed monthly data assimilation in the stratified summer period, whereas all data assimilation frequencies resulted in skillful forecasts across depths in the mixed spring/autumn periods for shorter forecast horizons. Our results suggest that lower-frequency data (i.e., weekly) may be adequate for developing accurate forecasts in some applications, further enabling the development of forecasts broadly across ecosystems and ecological variables without high-frequency sensor data.

Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir

Heather L. Wander[1]*, R. Quinn Thomas[1,2], Tadhg N. Moore[1,2], Mary E. Lofton[1], Adrienne Breef-Pilz[1], Cayelan C. Carey[1]

[1]Department of Biological Sciences, Virginia Tech, Blacksburg, Virginia, USA

[2]Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, Virginia, USA

*Corresponding author: hwander@vt.edu

Submitted as an Article to *Ecosphere:* Methods, Tools, and Technologies Track

Abstract:

Ecosystems around the globe are experiencing increased variability due to land use and climate change. In response, ecologists are increasingly using near-term, iterative ecological forecasts to predict how ecosystems will change in the future. To date, many near-term, iterative forecasting systems have been developed using high temporal frequency (minute to hourly resolution) data streams for assimilation. However, this approach may be cost-prohibitive or impossible for forecasting ecological variables that lack high-frequency sensors or have high data latency (i.e., a delay before data are available for modeling after collection). To explore the effects of data assimilation frequency on forecast skill, we developed water temperature forecasts for a eutrophic drinking water reservoir and conducted data assimilation experiments by selectively withholding observations to examine the effect of data availability on forecast accuracy. We used in-situ sensors, manually collected data, and a calibrated water quality ecosystem model driven by forecasted weather data to generate future water temperature forecasts using FLARE (Forecasting Lake And Reservoir Ecosystems), an open-source water quality forecasting system. We tested the effect of daily, weekly, fortnightly, and monthly data assimilation on the skill of 1 to 35-day-ahead water temperature forecasts. We found that forecast skill varied depending on the season, forecast horizon, depth, and data assimilation frequency, but overall forecast performance was high, with a mean 1-day-ahead forecast root mean square error (RMSE) of 0.94°C, mean 7-day RMSE of 1.33°C, and mean 35-day RMSE of 2.15°C. Aggregated across the year, daily data assimilation yielded the most skillful forecasts at 1-7-day-ahead horizons, but weekly data assimilation resulted in the most skillful forecasts at 8-35-day-ahead horizons. Within a year, daily to fortnightly data assimilation substantially outperformed monthly data assimilation in the stratified summer period, whereas all data assimilation frequencies resulted in skillful forecasts across depths in the mixed spring/autumn periods for shorter forecast horizons. Our results suggest that lower-frequency data (i.e., weekly) may be adequate for developing accurate forecasts in some applications, further enabling the development of forecasts broadly across ecosystems and ecological variables without high-frequency sensor data.

Key Words: data collection frequency; FLARE; high-frequency sensors; initial conditions; observations; uncertainty; water temperature

Introduction

In the face of increasing ecological variability due to climate and land use change (e.g., Gilarranz et al., 2022, Malhi et al., 2020), ecological forecasting is increasingly being used for understanding and predicting future ecological change (Carey et al., 2022d, Lewis et al., 2022). Here, we define ecological forecasts as predictions of future environmental conditions with quantified uncertainty (see Carey et al., 2022d, Lewis et al., 2022). Applications of ecological forecasts can improve understanding of ecosystem processes (e.g., carbon cycling, Bett et al., 2020), quantify predictability of environmental variables (e.g., rodent abundances, White et al., 2019), and inform management of ecosystem services (e.g., fisheries management, Lindegren et al., 2010). Because of their broad utility, forecasts are increasingly being developed by the research community to predict population, community, and ecosystem dynamics (Lewis et al., 2022). For example, an ongoing, community-based forecasting challenge organized by the Ecological Forecasting Initiative's Research Coordination Network has received thousands of ecological forecast submissions of National Ecological Observatory Network (NEON) data (e.g., lake water temperature, tick abundances, forest net ecosystem production, beetle communities) before the data have been collected (Thomas et al., 2023a).

Many near-term (daily to decadal) ecological forecasts are produced using the iterative, near-term forecasting cycle, in which models are updated as new observational data become available to generate forecasts into the future with quantified uncertainty (Dietze et al., 2018). The process of updating forecast models with newly available data, termed data assimilation (DA), is a critical component of the iterative, near-term forecast cycle (Dietze et al., 2018, Luo et al., 2011). DA allows for iterative updating of ecological hypotheses and models as forecasts are continuously assessed and updated with the most recent ecosystem observations (Dietze et al., 2018; White et al., 2019). DA can also improve forecast accuracy by updating forecast model initial conditions (i.e., starting values given to the model), states, and/or parameters at the timestep that the new observations become available (Cho et al., 2020, Gottwald and Reich, 2021, Luo et al., 2011, Niu et al., 2014). For example, McClure et al. (2021) developed 1 to 2-week-ahead forecasts of reservoir methane emissions with and without weekly DA and found that the accuracy of forecasts with DA was 44 - 128% higher than forecasts without DA over a five-month forecasting period. Despite the usefulness of DA for improving forecasts, however, the optimal frequency of observations for updating ecological models to produce skillful forecasts is not well characterized.

While there are a number of best practices proposed for applying the near-term, iterative forecast cycle in ecology (e.g., Clark et al., 2001, Harris et al., 2018, Lewis et al., 2022, White et al., 2019), few recommendations exist for choosing the optimal frequency of DA to produce accurate forecasts. Specifically, determining the appropriate frequency of observations for DA across a range of ecological variables is needed to improve the scalability of ecological forecasting, particularly if accurate forecasts can be developed using lower frequency observations. For example, if weekly or fortnightly DA yielded similarly accurate lake dissolved oxygen forecasts as daily DA, then water quality forecasting systems could be developed for lakes that have weekly or fortnightly routine monitoring program data without needing expensive high-frequency sensors, thereby enabling forecasts to be generated for many waterbodies globally.

Currently, many automated ecological forecasting systems rely on high-frequency sensors to assimilate data at each time step and generate accurate forecasts (e.g., Baracchini et al., 2020b, Corbari et al., 2019, Marj and Meijerink, 2011, Page et al., 2018, Tanut et al., 2021), but it is possible that high-frequency sensor data collection may not be needed for DA. Moreover, deployment of high-frequency sensors is not always feasible for all ecological variables (e.g., zooplankton abundance, biogeochemical concentrations, Marcé et al., 2016) and some remote locations have additional logistical constraints for maintaining autonomous sensor operation (Steere et al., 2000). Furthermore, some remotely sensed variables may only be available as satellite orbits and weather conditions (e.g., cloud cover) allow (e.g., Herrick et al., 2023). Thus, identifying how best to integrate observational data collected at different temporal frequencies into forecast models has emerged as a critical need for ecological forecasters (LaDeau et al., 2017).

Studies on the frequency of DA for environmental forecasts have generally shown that more temporally frequent DA improves forecast accuracy, but not always, which may be related to the sensitivity of forecasts to model initial conditions. For example, DA occurring every 24 hours using in-situ snow data (e.g., snow depth, density, snow water equivalent) resulted in better predictions of these snow variables in an alpine snowpack model compared to DA occurring every 3 hours (Piazzi et al., 2018). Conversely, DA 'experiments' performed for NOAA (National Oceanic and Atmospheric Administration)'s Global Forecasting System using meteorological observations collected at different frequencies showed that DA occurring every 2 hours resulted in more accurate air temperature and wind speed forecasts compared to DA occurring every 6 hours (He et al., 2020). These differences are likely because uncertainty in meteorological forecasts is primarily driven by the forecast model's initial conditions. Thus, more frequent DA, which constrains the model's initial conditions, will almost always improve the skill of meteorological forecasts (e.g., Clark et al., 2016, He et al., 2020, Simonin et al., 2017). In contrast, for forecasts of environmental systems in which model process uncertainty and model driver data uncertainty are more important sources of uncertainty (e.g., Dietze, 2017a, Heilman et al., 2022, Lofton et al., 2022, Thomas et al., 2020), it is unknown whether more frequent DA can improve forecast skill by generating initial conditions more consistent with observations.

To the best of our knowledge, there have been only a few ecological DA experiments that have tested the effects of different observation frequencies on forecast skill (e.g., Massoud et al., 2018, Piazzi et al., 2018, Weng and Luo, 2011, Ziliani et al., 2019), and none that have considered how the frequency of data used for assimilation affects forecast skill across both spatial and temporal scales. Weng & Luo (2011) assimilated eight different carbon datasets (e.g., root biomass, litter fall, soil respiration), each with different collection frequencies, to identify the relative importance of these data sources in constraining long-term carbon dynamics, but did not consider how different frequencies of the same dataset could affect forecast skill. Piazzi et al. (2018) assimilated multiple snow observations at two different frequencies (3 and 24 hours) for predicting different snow-related variables (e.g., depth, density, and snow water equivalent), and Ziliani et al. (2019) performed DA tests using 1-20 second assimilation of water depth data to assess water level forecast skill, but neither considered the effect of less frequent assimilation (e.g., >24 hours). Massoud et al. (2018) performed DA tests using a wider range of temporal frequencies (e.g., ~3-34-day abundance data) to predict plankton community dynamics, but did not consider the effects of DA across spatial scales (i.e., how DA affects forecast skill across multiple sites or depths in an aquatic ecosystem). As a result, further work is needed to quantify the utility of increased observation and DA frequency over both time and space to forecast performance in ecological systems with varying sensitivities to initial conditions.

Among ecosystems, freshwater lakes and reservoirs are particularly important systems for developing near-term forecasts because they provide essential ecosystem services, including drinking water, food, irrigation, and recreation (Carpenter et al., 2011, Meyer et al., 1999, Williamson et al., 2016). Because freshwaters are experiencing greater variability and adverse water quality issues in response to land use and climate change (e.g., O'Reilly et al., 2015, Paerl and Paul, 2012, Woolway et al., 2021), some water managers have used forecasts to preemptively address poor water quality events (reviewed by Lofton et al., 2023). To date, iterative, near-term freshwater forecasts have been developed for a number of water quality variables, including water temperature (e.g., Carey et al., 2022d, Thomas et al., 2023b), dissolved oxygen (e.g., Wang et al., 2016), and phytoplankton (e.g., Page et al., 2017, Woelmer et al., 2022). These forecasts have been developed using DA with observations collected by high-frequency sensors at intervals ranging from 4 minutes to 24 hours. However, most manual-sampling water quality monitoring programs collect observations on weekly to fortnightly scales (e.g., Francy et al., 2015, Kirchner and Neal, 2013, Romero et al., 2002), currently precluding the scaling of existing forecasting systems broadly and underscoring the need to determine whether less frequent observations can be used to produce accurate forecasts.

To quantify how DA at different frequencies affects forecast skill up to 35 days into the future, we performed DA experiments in which we separately assimilated daily, weekly, fortnightly, and monthly data into reservoir water temperature forecasts. Water temperature forecasts are used to inform management decisions on water extraction depth and preemptive water quality interventions (Georgakakos et al., 2005; Kehoe et al., 2015; Mi et al., 2020), and thus our study has much utility for both informing how best to forecast complex ecosystem dynamics, as well as manage drinking water supplies. Our research questions were: 1) Which frequency of DA generates the most skillful water temperature forecasts? 2) How does forecast skill vary across time (specifically focusing on the mixed vs. stratified seasons within a year) and space (i.e., reservoir depth)? and 3) How does DA frequency influence total forecast uncertainty and what is the relative contribution of initial condition uncertainty to total forecast uncertainty? As previous work has suggested that reservoir water temperature forecasts can sometimes exhibit sensitivity to initial conditions (Thomas et al. 2020), we expected that less frequent DA would result in decreased forecast skill and increased total uncertainty. In addition, we expected that forecast skill would be better at deeper depths, especially during thermally-stratified periods (e.g., Mercado-Bettín et al. 2021; Thomas et al. 2020).

Methods

Forecasting system overview

We applied the Forecasting Lake And Reservoir Ecosystems (FLARE) forecasting system (Thomas et al., 2020) to Beaverdam Reservoir, Virginia, USA to produce daily water temperature forecasts for 1-35 days into the future (hereafter referred to as forecast horizon) during 1 January 2021 - 31 December 2021. FLARE is an open-source forecasting system that incorporates real-time water quality sensor data, DA, ensemble-based forecasts, and uncertainty quantification to predict near-term water quality conditions (Thomas et al., 2020).

Forecast generation via FLARE can be summarized by four steps (Figure 1). First, 10-min resolution water temperature data were collected by sensors deployed in the reservoir (Figure 1 step 1). Second, these data were transferred to the cloud and stored in a GitHub repository, where they were downloaded daily and made available for DA (Figure 1 step 2). Simultaneously, 1 to 35-day-ahead NOAA meteorological forecasts were downloaded daily as driver data for the reservoir hydrodynamic model to generate the water temperature forecasts. Third, during the forecast generation step, DA was used to update initial conditions and parameters with the most recent observations using an ensemble Kalman filter, a numerical approach that allows for the updating of model states and parameters using data (Evensen, 2003) (Figure 1 step 3a). Following DA, the reservoir hydrodynamic model was initialized with the updated model states and parameters to produce 1-35-day-ahead forecasts for each 0.5 m depth interval across the water column (Figure 1 step 3b). Finally, forecast skill was assessed by comparing observed vs. predicted water temperatures for each daily forecast at each depth (Figure 1 step 4). We repeated steps 3a-4 for daily, weekly, fortnightly, and monthly intervals of DA throughout the year as part of the DA experiments to compare forecast skill over time.

Study site and monitoring

Beaverdam Reservoir (BVR) is a small (0.28 km^2), shallow (Z_{max} = 11 m), dimictic, eutrophic reservoir in southwestern Virginia, USA (37.31° N, 79.82° W; Figure 2). BVR is managed by the Western Virginia Water Authority as a secondary drinking water supply and is located in a deciduous forest catchment (Doubek et al., 2019). During a typical year, BVR is stratified from mid-March to late October and mixed from November to early March (Hounshell et al., 2021). BVR experiences summer hypolimnetic anoxia and cyanobacterial blooms, both of which are controlled by water temperature and thermal stratification (Doubek et al., 2019; Hamre et al., 2018), making forecasts of water temperature important for water quality management.

Water quality monitoring of BVR includes both manual sampling and high-frequency sensors. From 2014-present, manual water quality sampling occurred weekly to fortnightly during the summer stratified period and fortnightly to monthly during the remainder of the year (Carey et al., 2022c). Starting in June 2020, high-frequency sensors were deployed in the reservoir, enabling a range of DA frequencies to be compared in this study. We deployed NexSens T-Node FR Temperature Sensors (NexSens Technology, Fairborn, OH, USA) at 1 m intervals from the surface to sediments and a YSI EXO2 sonde (YSI Incorporated, Yellow Springs, OH, USA) that monitored temperature at 1.5 m at the deepest site in BVR (Figure 1; see Carey et al., 2023 for sensor information). These sensors collected data every 10 minutes, which was transmitted every 3 to 9 hours via secure sensor gateways to a Git repository in the cloud (Carey et al., 2023, Daneshmand et al., 2021). We removed observations collected during periods of sensor maintenance, as well as depth-adjusted the data using an offset calculated from a CS451 Stainless-Steel Pressure Transducer (Campbell Scientific, Logan, UT, USA) to account for water level changes (Wander et al., 2023b). Because of this range in latency, or the time that it takes for data to become available for modeling after they are initially collected, we used the daily mean in our forecasting application. Following quality checks, these data were integrated into the FLARE forecasting system to produce depth-specific daily water temperature forecasts.

Hydrodynamic model configuration

For modeling reservoir hydrodynamics, we used the General Lake Model (GLM) v.3.3.0 (Hipsey et al., 2022) to forecast water temperature in BVR. GLM is an open source, 1-D process-based hydrodynamic model commonly used within the freshwater research community to simulate water quality in lakes and reservoirs (Hipsey et al., 2019). GLM uses a Lagrangian approach for simulating different water layers and has been applied to a variety of lakes worldwide for modeling (e.g., Bruce et al., 2018, Read et al., 2014) and forecasting hydrodynamics (e.g., Thomas et al., 2020, 2023b).

We configured GLM for BVR using historical bathymetric data (Carey et al., 2022b) and water temperature observations for initial conditions (Carey et al., 2023). We configured GLM with two sediment zones to simulate epilimnetic (surface) and hypolimnetic (bottom) sediment temperature dynamics following Carey et al. (2022a). GLM requires meteorological and reservoir inflow observations as driver data to run the model. Because we were applying GLM for forecasting, meteorological forecasts, not observed meteorology, were used as driver data in the model, as described below. Additionally, we set the inflow to equal outflow in this study given limited inflow data for validation and the relatively short forecast horizons (≤ 35 days). We initiated the model using its default parameter set (Hipsey et al., 2019) and performed calibration via a 35-day spin-up period with DA to tune parameters before the start of our focal forecasting period (described below).

FLARE configuration for DA and uncertainty

We configured FLARE for BVR following its application to other lakes and reservoirs (Thomas et al., 2020, 2023b). We set the number of forecast ensemble members to 256 to ensure an adequate representation of uncertainty and prevent the ensemble Kalman filter from developing erroneous correlations among ensemble members that can occur with low ensemble sizes (Duc et al., 2021, Machete and Smith, 2016). While we used default values for most GLM parameters, we used the ensemble Kalman filter in FLARE to tune three model parameters that we identified as important for water temperature simulations using GLM in a similar, nearby reservoir (Carey et al., 2022a, Thomas et al., 2020): 1) the longwave radiation scaling factor (hereafter, *longwave*); 2) epilimnetic sediment temperature parameter (hereafter, *epi_sed_temp*); and 3) hypolimnetic sediment temperature parameter (hereafter, *hypo_sed_temp*).

We used state augmentation to tune the three parameters in the ensemble Kalman filter (Thomas et al. 2020). Specifically, correlations between the parameter values and the model states with observations (i.e., water temperatures at the depths with sensor observations) were used to adjust parameters to be consistent with the most recent data used in DA. The three tuned parameters were initially calibrated during a spin-up period from 27 November to 31 December 2020 and were subsequently updated via DA throughout the forecasting period. To avoid the common issue of artificially low parameter uncertainty in sequential DA (Dietze, 2017a), we specified the standard deviation of a normal distribution for each parameter (1.0°C for the sediment temperature parameters and 0.02 for the longwave radiation scaling factor). Initial exploration of parameter fitting in this study indicated that the application of FLARE over the full year resulted in low parameter uncertainty, necessitating us to specify the standard deviation a priori rather than estimating it using DA. The distributions we chose were adapted from a prior application of FLARE that estimated the standard deviation of parameter distributions across six lakes (Thomas et al., 2023b).

FLARE uses a numerical ensemble-based approach to simulate and propagate forecast uncertainty (Thomas et al., 2020). We represented the contribution of uncertainty from meteorological driver data, initial conditions, model process, and model parameters using the 256-member ensemble, following Thomas et al. (2020). First, to represent the contribution of meteorological driver data uncertainty, we assigned each of the 256 FLARE ensemble members one of the 30 ensemble members from the 1-35-day-ahead meteorological forecasts (National Oceanic and Atmospheric Administration's Global Ensemble Forecasting System) to drive GLM for forecasting. Second, we represented uncertainty in the initial conditions of the forecasts using the spread in model states among the 256 ensemble members on the first day of each forecast. This spread was determined by either using the prior day's forecast as a starting point for the next day's forecast (when no data were available for DA) or the updated states following DA (when data were available for DA). We set the observation uncertainty standard deviation to

0.1°C, determined from the standard deviation of temperature observations and following prior applications of FLARE (Thomas et al. 2020). Third, we represented model process uncertainty by adding random noise to the water temperature predictions from each of the 256 FLARE ensemble members at each daily time-step in a 1-35-day-ahead forecast horizon. The random noise for each modeled depth within an ensemble member was drawn from a normal distribution with a standard deviation of 0.75°C, as used in a previous application of FLARE that reported well-calibrated forecast uncertainty (Thomas et al. 2020). The random noise was spatially correlated so that it was most similar for nearby depths and most different for further-apart depths. The strength of the spatial correlation was determined by the exponential decay of the correlation strength with distance (Thomas et al., 2023b). Fourth, we represented parameter uncertainty using the standard deviations of the distributions for the three tuned GLM parameters described above. A unique parameter value drawn from each of the three distributions was assigned to each of the 256 FLARE ensemble members. The parameter value assigned to an ensemble member was only updated when DA occurred. Parameters not tuned by the ensemble Kalman filter were assumed to have fixed values and uncertainty in these parameters was not calculated.

To determine whether there was a relationship between the magnitude of initial conditions uncertainty and the sensitivity of forecast skill to more frequent DA (following Clark et al., 2016, He et al., 2020, Simonin et al., 2017), we quantified the contribution of initial conditions uncertainty to total forecast uncertainty in our DA forecasts for all DA frequencies. For this analysis, we isolated the magnitude of initial conditions uncertainty by generating the water temperature forecasts for all 365 days with and without initial conditions uncertainty and compared the variance among all 256 ensemble members. We also calculated the proportion of initial conditions uncertainty within total forecast uncertainty for all depths, horizons, and stratified vs. mixed periods.

Data assimilation experiments

To quantify the effect of DA at different frequencies on forecast skill, we conducted DA experiments in BVR from 1 January to 31 December 2021 (n = 365 days). As noted above, we used a spin-up period from 27 November - 31 December 2020 (n = 35 days) during which DA occurred, but no forecasts were generated. During the one-year forecast period in 2021, we forecasted daily water temperature at 23 depths in the reservoir (spanning 0.1 to 11 m depth at 0.5 m intervals) and assessed forecast performance relative to observations across each forecast's daily predictions for 1 to 35-day-ahead horizons and depth intervals. We focused on three focal depths (1, 5, and 9 m) when reporting results, as these depths are representative of the surface, middle, and bottom layers of the water column, respectively. We chose 9 m to represent the bottom of the reservoir because deeper depths were not always observed due to variability in water levels throughout the year (within ~1 m due to seasonality in flows).

We performed DA experiments using four different DA frequencies (daily, weekly, fortnightly, and monthly) to represent different data collection latencies that are commonly used by water quality monitoring programs (e.g., Engelhardt and Kirillin, 2014, Francy et al., 2015, Kirchner and Neal, 2013, Liu et al., 2019, Romero et al., 2002). We assimilated water temperature data across different temporal frequencies by downsampling from the high-frequency observations collected by our sensors. This resulted in four different temporal frequencies for which DA occurred, corresponding to either daily (representing standard FLARE DA), weekly, fortnightly, or monthly DA (see Data Assimilation Experiments box in Figure 1 and Appendix S1: Figure S1 for visualization of DA frequencies). For example, for the weekly DA frequency, observations were selected every seven days starting on 4 January 2021 and ending on 31 December 2021. In this example, DA only occurred once per week; the forecasts that were generated on the six other days in the same week did not include DA (i.e., no DA occurred during 5 January - 10 December 2021 even though forecasts were still generated daily during this interval; Figure 1). Fortnightly and monthly DA occurred every 14 days and 30 days, respectively, throughout the year.

We generated 365 daily forecasts starting on 1 January 2021 for each of the four DA frequencies. While we recognize that we are producing hindcasts for a historical period, because the model was forced with only forecasted drivers and out-of-sample forecast evaluation occurred, we refer to these retrospective forecasts or hindcasts as forecasts throughout for consistency (following Jolliffe and Stephenson, 2012).

Analysis

Question 1: For all n = 1460 forecasts produced (365 forecasts generated daily over a year for four different DA frequencies), we used root mean square error (RMSE) and continuous ranked probability score (CRPS; Gneiting et al., 2005) to quantify forecast skill. We defined skillful water temperature forecasts as those with an RMSE < 2°C, a commonly-used threshold for lake and reservoir hydrodynamic modeling following Bruce et al. (2018), Read et al. (2014), and many others. Mean full water column RMSE was calculated for each of the 35 days across all forecast horizons for each DA frequency regardless of whether data were assimilated the day the forecast was generated. We aggregated RMSE across depths and dates to determine the lowest temporal frequency of DA required to generate the most skillful water temperature forecasts across the full water column and throughout the entire forecast period. We focus on RMSE in the results and all CRPS values are reported in the SI.

Question 2: Using RMSE and CRPS, we compared forecast skill across depths and seasons to identify how the frequency of DA affected forecast accuracy over space and time. To quantify spatial forecast performance, we calculated RMSE and CRPS for each depth (1-11 m) at each forecast horizon (1-35 days ahead) and DA frequency in BVR. To quantify temporal forecast performance, we compared forecast skill at each horizon aggregated within thermally-stratified vs. mixed periods in BVR. The stratified period began on the first day that the water density difference between the reservoir surface (0.1 m) and the maximum depth observed for the reservoir on each day (e.g., between 9-11 m) was ≥ 0.1 kg/m^3 for at least three consecutive days (following Ladwig et al., 2021). Conversely, the mixed period began on the first day that surface and bottom water density differences were < 0.1 kg/m^3 for at least three consecutive days. Altogether, we compared forecast skill between stratified vs. mixed periods; among depths (1, 5, and 9 m), and among forecast horizons (focusing in on 1, 7, and 35-day-ahead forecasts) for each of the four DA frequencies.

Question 3: We quantified total forecast uncertainty for each day in the 1-35-day forecast horizon using the variance of the 256-member FLARE ensemble. The relative contribution of initial condition uncertainty to total forecast uncertainty was calculated for each forecasted day by comparing the variance in the 256-member FLARE ensemble between the set of forecasts with initial condition uncertainty included and the set without initial condition uncertainty.

All statistical analyses were conducted in R v.4.2.0 (R Core Team, 2022). All R code and data files used to run these analyses are archived and available in the Zenodo repository (Wander et al., 2023a, 2023b).

Results:

BVR water temperature dynamics

BVR exhibited typical annual water temperature dynamics during the forecasting period in 2021. Water temperature throughout the water column ranged from 1.4 to 29.9°C during the year. The summer stratified period began on 12 March and ended on 7 November 2021, and the reservoir was mixed from 1 January - 11 March and 8 November - 31 December (Figure 3). Thermocline deepening occurred throughout the summer stratified period, starting at 1.5 m in March with stratification onset and deepening to 9.5 m in November before fall turnover (Figure 3). During the winter, there were three brief periods of ice cover of one to three days in duration in January and February when inverse stratification occurred (Figure 3; Carey and Breef-Pilz, 2022). We removed these few ice-cover days from the analysis and grouped mixed (n = 118 days) vs. summer stratified data (n = 241 days) for analysis.

Data assimilation frequency altered forecast output and parameters over time

We were able to successfully forecast water temperature throughout the water column over the year using DA to update model states and parameters (Figures 4-5). Across all depths, DA constrained uncertainty by updating initial conditions with the most recent water temperature observations. Forecast uncertainty for the lower DA frequencies was strongly dependent on the time since last assimilation (Figure 4). On average, forecast variance at the one-day horizon across 2021 for forecasts with daily DA was 1.56°C while mean forecast variance at the one-day horizon for forecasts with monthly DA was 3.25°C.

We observed that DA frequency altered parameter evolution of the forecasts (Figure 5). The daily DA frequency resulted in more variable parameter estimates through time for all three tuned parameters, reflecting the more frequent adjustment that occurred each time data were assimilated. Importantly, parameter evolution for forecasts with daily DA yielded very different estimates than the weekly, fortnightly, and monthly DA forecast frequencies (Figure 5). For example, the evolution of the longwave radiation scaling parameter (longwave) over the 365-day forecast period showed that forecasts with weekly, fortnightly, and monthly DA frequencies converged at ~0.91 by December 2021, whereas the longwave parameter for forecasts with daily DA was at ~0.85 by the end of the year (Figure 5a). Similarly, the parameter controlling the surface layer sediment temperature (epi_sed_temp) in daily DA forecasts began to diverge from the other DA frequencies in early April (Figure 5c). The non-daily DA frequencies (i.e., weekly, fortnightly, monthly DA) surface sediment layer temperature parameter (epi_sed_temp) values ranged from 13.51 to 15.62°C, whereas the daily DA frequency epi_sed_temp values ranged from 13.29°C to 17.0°C during April-December. For the parameter controlling the bottom layer sediment (hypo_sed_temp), daily DA forecasts exhibited much more variable values (ranging from 10.24°C to 11.21°C) than forecasts for any other DA frequency (range 10.65°C to 10.72°C; Fig 5b) from April to December.

Question 1: Which frequency of data assimilation generates the most skillful water temperature forecasts?

Aggregated among depths and time periods, weekly DA resulted in the most skillful water temperature forecasts of the four DA frequencies for the greatest number of 1-35-day-ahead horizons (Figure 6). Among horizons, we observed that the frequency of DA needed to produce skillful forecasts varied (Figure 6). At shorter horizons (1-7 days ahead), daily DA resulted in the most skilled forecasts, but at longer horizons (8-35 days ahead), weekly DA resulted in the most skilled forecasts (Figure 6).

The skill of all forecasts degraded as the forecast horizon increased, but the decrease in performance was greatest for daily DA forecasts, such that forecasts generated using monthly, fortnightly, and weekly DA all outperformed daily DA forecasts by the 19-day forecast horizon (Figure 6), when aggregating across all depths and time periods. The daily DA forecasts exceeded the 2°C RMSE metric of skill on the 28-day-ahead horizon, whereas the weekly, fortnightly, and monthly forecasts never exceeded that metric for any of the 1-35-day-ahead horizons. These results were consistent across forecast evaluation metrics, including the CRPS metric that evaluates the full ensemble forecast (Appendix S1: Figure S2).

Question 2: How does forecast skill vary across time and space?

Aggregated across depths, horizons, and DA frequencies over the year, forecast skill overall was high, with a mean water temperature forecast RMSE of 1.53 ± 1.86°C (1 S.D.). Forecast skill was generally best at 9 m regardless of horizon or DA frequency. Aggregated 9 m forecast skill was 1.29 ± 1.80°C, followed by aggregated 5 m forecast skill (1.63 ± 1.85°C), and aggregated 1 m forecast skill (1.69 ± 1.82°C). As expected, forecast skill generally decreased with horizon, with a mean 1-day-ahead forecast RMSE of 0.80 ± 1.20°C, mean 7-day RMSE of 1.15 ± 1.60°C, and mean 35-day RMSE of 1.99 ± 2.17°C. However, we observed an exception to this pattern for 1 m mixed forecasts, which is further described below.

On average, forecast skill was slightly better (as indicated by smaller RMSE) during the stratified period than during the mixed period, aggregated among all depths and horizon regardless of DA frequency (aggregated mixed RMSE = 1.59 ± 1.57°C, stratified RMSE = 1.46

± 2.13°C; Figure 7). Forecast skill was more variable among forecast horizons than depths in the mixed period, whereas forecast skill was variable across both depths and horizons in the stratified period (Figure 7). In the stratified period, forecast skill was best at 9 m, with relatively similar skill over the forecast horizon (Figure 7f). In the mixed period, forecast skill varied very little among depths aggregated across horizons (Figure 7a, c, e), with consistently greater decreases in skill with increasing horizon than in the stratified period, except for at 1 m. Forecast skill at 1 m decreased rapidly until ~the 19-day horizon, after which forecast skill remained constant for the daily DA and increased for the weekly, fortnightly, and monthly DA frequencies until the end of the forecast horizon (Figure 7a).

While daily DA always resulted in the best forecast skill for 1-day-ahead horizons, lower frequency DA typically outperformed daily DA as the forecast horizon increased. An exception was for 9 m stratified forecasts, when daily DA resulted in the lowest RMSE for all forecast horizons and never exceeded 1.51°C for the duration of the 35-day forecast horizon (Figure 7f). Additionally, 9 m stratified forecasts were the only forecasts where skillful (RMSE < 2°C) forecasts were produced for all DA frequencies and horizons (Figure 7f).

Question 3. How does DA frequency influence total forecast uncertainty and what is the relative contribution of initial condition uncertainty to total forecast uncertainty?

Lower frequency DA forecasts consistently had more total uncertainty (Figure 8). We found that the differences between uncertainty for daily and monthly DA were largest at 1-day-ahead horizons and largely converged by the end of the 35-day horizon (Figure 8). At 1 m depth, total uncertainty was similar between the mixed and stratified periods across the 35-day horizon, but at 5 and 9 m, total uncertainty was on average higher in the stratified than mixed period. Both

RMSE and total variance were similar for forecasts run with and without initial conditions uncertainty included (Appendix S1: Figures S3-S4).

Forecasts with less frequent DA had a greater contribution of initial condition uncertainty to total forecast uncertainty during the first few days of the forecast horizon. However, overall, initial conditions uncertainty contributed a minimal proportion of the total uncertainty for forecasts generated with daily DA (Figure 9). At the 1-day-ahead forecast horizon, daily DA initial conditions uncertainty was 0% of total uncertainty, whereas initial conditions uncertainty contributed 55 - 71% of total forecast uncertainty in forecasts for all other DA frequencies (Figure 9). The role of initial conditions uncertainty for all depths in the mixed period and surface forecasts in the stratified period was minimal (< 1%) across all DA frequencies after the 10-day horizon (Figure 9a-c, e). Conversely, initial conditions uncertainty made up a larger proportion of total forecast uncertainty for stratified 5 m and 9 m forecasts for forecast horizons between 10 and 20 days (ranging from 5-10%; Figure 9d, f).

Discussion:

Across a year of water temperature forecasts in our focal reservoir, we found that weekly DA generally resulted in the most skillful water temperature forecasts. However, skill varied among depths, forecast horizons, and time of year, suggesting that DA frequency should be chosen based on the specific forecast application. For example, if water temperature forecasts are specifically needed to guide decision-making that involves the deeper reservoir layers (e.g., 5 m or 9 m) at short time horizons (e.g., <5 days ahead), daily DA might be most advantageous (Figure 7). Conversely, if water temperature forecasts are needed for the surface water at 20-35 day-ahead horizons, then weekly to monthly DA may be sufficient (Figure 7). Despite the usefulness of DA for improving forecast skill, more frequent DA did not always lead to more skillful water temperature forecasts, in part because initial conditions uncertainty only comprised a significant proportion of total forecast uncertainty within the first few days of the forecast horizon (Figure 9). Below, we interpret our results for each research question and make recommendations for considering which DA frequency might be appropriate for different ecological forecast applications.

Q1: Which frequency of data assimilation generates the most skillful water temperature forecasts?

In this study we found that less frequent DA (e.g., weekly, fortnightly, and monthly DA) sometimes led to more skillful water temperature forecasts than daily DA for all depths during the mixed period. This pattern of weekly DA outperforming daily DA forecast skill during the mixed period is likely because daily DA led to parameter overfitting, as indicated by the greater short-term variability in parameter estimates over time (Figure 5). Because water temperatures are fairly stable at deeper depths, and thus daily observations can consistently predict tomorrow's water temperature accurately, parameter overfitting was less problematic for daily DA at hypolimnetic depths (Figure 7f). As a result, hypolimnetic forecast skill was best with daily DA during stratified conditions, but this pattern did not extend to other depths or the mixed period (Figure 7).

Our work is consistent with studies that have found that the optimal DA frequency often matches that of the forecast model timestep (e.g., Derot et al., 2020, Woelmer et al., 2022). For example, during both the mixed and stratified periods, daily DA was always better for 1-day-ahead forecasts, but was often outperformed by weekly DA at 8-day-ahead forecast horizons

(Figure 6). Because water temperatures were homogenous among all depths during the mixed period, water temperature variability among all depths was likely driven by air temperature variability, ultimately making it more challenging to predict water temperature across depths as the forecast horizon increased. During the stratified period, however, less frequent DA could still generate accurate surface and mid-depth water temperature forecasts. The increased importance of daily DA at bottom depths during the stratified period is likely because of the increased thermal stability at bottom depths associated with thermal stratification (Figure 7). This pattern is in contrast with other water temperature forecasting studies that have found daily DA necessary for improving the skill of forecasts in the middle of the water column around the thermocline (Baracchini et al., 2020a), but is likely explained by the overfitting of both the daily longwave radiation and the epilimnetic sediment temperature parameters (Figure 5).

We note that there are many ways to quantify skill beyond the 2°C RMSE threshold used here. We chose to use RMSE because it is a commonly used metric by lake modelers to determine the deviation between observed vs. modeled values (Bruce et al., 2018, Read et al., 2014). However, forecast skill could also be quantified by other metrics, such as the correlation coefficient, Nash–Sutcliffe model efficiency coefficient, percent relative error, normalized mean absolute error, or others (Bennett et al., 2013). While 2°C RMSE is a subjective criterion of forecast skill, we note that CRPS results followed similar patterns as our RMSE metric (Appendix S1: Figure S2), further supporting our results when evaluating the full distribution of the forecasts.

Q2: How does forecast skill vary across time and space?

Our use of a 35-day forecast horizon allowed us to compare water temperature predictability across multiple horizons at different depths and times of year, thereby elucidating patterns in ecosystem predictability across both space and time. We generally observed expected declines in forecast skill with increasing horizon, as noted in many other studies. However, 1-m forecast skill during the mixed period increased with lower frequency DA (weekly, fortnightly, and monthly), while forecast skill with daily DA leveled off at 19-35-day horizons (Figure 7a). Improved forecast skill for lower frequency DA suggests that our forecasts are capturing surface water temperature dynamics in the mixed period at longer horizons better than other depths, particularly those during the stratified period. This may be due to the smaller range in water temperature variation that occurs in the mixed period relative to the stratified period over a 35-day interval, allowing variance to level off at longer forecast horizons as water temperature observations better matched predicted values (Figures 7-8). While improved forecast skill at longer horizons has been observed in the literature (e.g., Wheeler et al., 2023), this pattern is often associated with variables that have predictable, cyclical patterns at long horizons (e.g., annual tree leaf-out).

Overall, we observed generally high forecast skill across all depths and times of year for most forecast horizons. Across DA frequencies, depths, and times of year, RMSE was only consistently above the 2°C threshold for daily DA at 28-35-day horizons (Figure 6). By the end of the 35-day forecast horizon, daily DA forecast skill for most depths and times of the year was >2°C, except 9 m stratified forecasts, which had a mean RMSE of 1.29 ± 1.8°C across DA frequencies. The higher forecast skill at 9 m is likely because fluctuations in bottom water temperatures were minimal during stratification (Figure 3).

Our findings are similar to other water temperature lake and reservoir forecasting studies. First, the pattern of increased forecast skill in the bottom waters is consistent with Mercado-

Bettín et al. (2021) and Thomas et al. (2020), who both found that the bottom water forecasts were more skillful than surface water forecasts. This is likely because bottom waters are not changing as much as surface waters throughout the year due to less atmospheric exchange. However, Clayer et al. (2023) found that surface water temperatures were more accurately simulated than bottom water temperatures, suggesting that the complex lake characteristics that control bottom water temperatures were not captured as well as the air temperature dynamics controlling surface water temperatures. Second, our finding that forecast skill was greater in the stratified period rather than mixed period is similar to the results of Thomas et al. (2020), likely due to the fact that water temperature dynamics were changing less among depths in stratified than mixed periods (Figure 3). Because of the variability in water temperature dynamics among seasons and depths, determining the conditions in which we can most accurately forecast water temperature can improve our understanding of ecosystem processes and functioning. Moreover, accurately forecasting water temperature is critical for forecasting additional lake and reservoir variables that are strongly driven by water temperature, such as phytoplankton biomass, dissolved oxygen concentrations, and greenhouse gas emissions (e.g., McClure et al. 2021).

Q3: How does DA frequency influence forecast uncertainty?

We found that initial conditions uncertainty contributed a substantial proportion of total uncertainty for weekly, fortnightly, and monthly DA, but only during the first few days of the forecast horizon. From 6-23 day-ahead horizons, the contribution of initial conditions decreased to <1% across all DA frequencies, depths, and times of year (Figure 9). We observed that high-frequency DA was required for skillful 9 m stratified forecasts, while weekly DA was sufficient for other depths and times. This finding may be because the contribution of initial conditions uncertainty decreases more rapidly within the first few days of the forecast horizon for the daily DA forecasts at 9 m in the stratified period. For all other depths and times of the year, the rate at which initial conditions uncertainty decreases is greater for weekly, fortnightly, and monthly DA, resulting in more similar performance of daily and weekly DA early in the forecast horizon (Figure 9). However, more frequent DA may not always improve forecast performance, especially when initial conditions uncertainty is not the dominant source of uncertainty, as seen at longer horizons. Given that initial conditions uncertainty predominated at the beginning of the forecast horizon, it is likely that total forecast uncertainty at longer horizons was primarily influenced by uncertainty in model process, model parameters, and/or meteorological driver data (Figure 9). Conversely, the dominant source of uncertainty for weather forecasting is typically initial conditions uncertainty given the inherent instability of atmospheric processes (Dietze, 2017b), which is why more frequent DA often substantially improves meteorological forecast skill.

Other lake and reservoir water quality forecasting studies have found that model driver data and process uncertainty were the dominant sources of total forecast uncertainty (Lofton et al., 2022, McClure et al., 2021, Thomas et al., 2020). Therefore, constraining other sources of uncertainty by using an ensemble approach or different forecasting models would likely further improve water temperature forecast skill. Additionally, using a different DA technique that uses a Bayesian approach to estimate a posterior distribution, rather than assuming that the parameters and model states are normally distributed, may also reduce uncertainty (e.g., particle filter; Wang et al., 2023). Because the dominant source of uncertainty in ecological forecasts will likely differ depending on the variable being forecasted, different DA techniques may not improve forecast skill equally among all ecological variables.

Recommendations for setting up DA for other forecasting systems

Determining whether an ecological forecasting application requires high-frequency sensors is necessary for increasing the scalability of ecological forecasting across ecosystems and variables. While high-frequency sensor data may improve forecast skill in some cases, sensor deployment is often costly, which limits the application of high-frequency data in some forecasting systems. Moreover, even if high-frequency sensors are deployed, identifying the minimum frequency of data required to make skillful ecological forecasts can be a useful exercise because high-frequency sensors malfunction and require maintenance, which can result in data gaps (e.g., Herrick et al., 2023). Many water quality forecasting applications to date have relied on high-frequency sensor data for assimilation to produce skillful forecasts of different variables (Cho and Park, 2019, Derot et al., 2020, Page et al., 2018). In this study, we found that daily DA only produced the most skillful 9 m stratified period water temperature forecasts, whereas weekly DA generally produced the most skillful surface and middle layer water temperature forecasts (Figure 7). Our findings indicate that high-frequency sensors may not be needed for accurate mixed period water temperature forecasts or surface layer forecasts in the stratified period.

The minimum frequency of DA needed to set up fully operational forecasting systems is likely to vary based on the ecosystem or forecast variable of interest. Depending on the water quality forecast application, different frequencies of data collection may be necessary to fully understand and predict water quality dynamics over time. For example, George and Hurley (2004) found that fortnightly observations were required to discern gradual trends in phytoplankton productivity, but monthly data were adequate for capturing declines in phytoplankton biomass over a 30-year period. Despite many successful applications of high-frequency DA in the literature for forecasting (e.g., Cho et al., 2020, Gottwald and Reich, 2021, Luo et al., 2011, Niu et al., 2014), not all ecological variables benefit from frequent DA, as not all variables are similarly forecastable.

In addition to the frequency of data collection, data latency can also affect the frequency of DA. Even for forecasting systems with high-frequency sensor data, data latency may reduce forecast skill if data are not immediately transmitted to forecasting workflows (e.g., they require a manual download) (Dietze et al., 2018). In cases with high data latency of the forecast variable (e.g., microscope counts of phytoplankton requiring laboratory analysis), data fusion approaches that assimilate multiple data sources may improve forecast skill (e.g., Baracchini et al., 2020b, Chen et al., 2021). For example, some studies have assimilated both in-situ measurements and remote sensing data to forecast reservoir water quality variables, including chlorophyll *a* and conductivity (Abdul Wahid and Arunbabu, 2022, Chen et al., 2021).

Finally, understanding the contributions of different sources of uncertainty can be useful for determining the DA frequency that generates the most skillful forecasts. Specifically, knowing the relative contribution of initial conditions uncertainty can inform sampling frequency needed to improve ecological forecast skill. For forecasts with total uncertainty dominated by process, parameter, or driver uncertainty, improving forecast skill may require modifying processes used for forecasting the ecological variable of interest, further constraining parameters by collecting more data, or improving weather forecast driver data (e.g., Grönquist et al., 2021).

Study Limitations

Our results suggest that weekly DA may suffice for some lake and reservoir water temperature forecasting applications, with the caveat that more frequent DA often improved water temperature forecast performance at short forecast horizons. However, we only assessed forecast skill for a single reservoir and ecological variable for only one year, and therefore note the limitations of extending these results to other systems and variables. Additionally, updating model parameters and initial conditions too regularly can lead to overprediction biases when forecasting, which may explain why weekly rather than daily DA resulted in more skillful water temperature forecasts in the mixed period (see Lin et al., 2021). Finally, because we did not quantify the contribution of all sources of uncertainty, we can only identify the relative role that DA has on reducing initial conditions uncertainty. Future studies that consider the role of other sources of uncertainty will improve our understanding of DA on total forecast uncertainty.

Conclusions

This study emphasizes the importance of DA for improving ecological forecast skill and has implications for forecasting efforts among a wide range of ecosystems and ecological variables. We argue that weekly observations of water temperature are likely "good enough" to set up a skillful forecasting system for many reservoir management applications, while daily DA would be most useful for applications requiring high forecast accuracy in the bottom waters or at short (< 5 - 7 day) forecast horizons. Because water temperature dynamics control many biological, chemical, and physical lake processes (Magnuson et al., 1979, Read et al., 2019, Yvon-Durocher et al., 2012), water temperature must be accurately forecasted before we can forecast other water quality variables. Therefore, determining ways to improve water temperature forecasts will have broad utility for advancing the development of many additional water quality forecasting systems.

Because near-term, iterative forecasts are particularly well suited to address ecological questions (Carey et al., 2022d, Dietze et al., 2018, White et al., 2019), determining how best to design and deploy ecological near-term, iterative forecasting systems is a pressing need (Diez et al., 2012, Ibáñez et al., 2013, Moustahfid et al., 2021). With the increasing deployment of high-frequency sensor networks (e.g., National Ecological Observatory Network (NEON) and Global Lake Ecological Observatory Network (GLEON); Mantovani et al., 2020, Marcé et al., 2016, Park et al., 2020) comes a growing need to understand how best to use these sensor data for forecasting. In response, we advocate for using DA experiments across ecosystems and ecological variables to determine how best to integrate observational data into iterative forecasting systems.

Conflict of Interest Statement: The authors declare no conflicts of interest.

Data Availability Statement: All data (Wander et al., 2023a) and code (Wander et al., 2023b) are publicly available on Zenodo (https://doi.org/10.5281/zenodo.7951402; https://doi.org/10.5281/zenodo.7958471)

Acknowledgements: This work was supported by the U.S. National Science Foundation (DEB-1753639, DBI-1933016, DBI-1933102, DEB-1926050, DEB-1926388, and OAC-2004323). We thank the Western Virginia Water Authority for their long-term access to Beaverdam Reservoir and the CIBR-FLARE team, Carey Lab, and Reservoir Group for helpful feedback throughout this project.

Author Contributions: HLW, RQT, and CCC co-developed the design of this study. RQT developed the FLARE framework for data assimilation experiments and forecasting workflow used in this study. TNM helped develop early iterations of the forecasting workflow. ABP oversaw sensor data collection. HLW wrote the initial draft of the manuscript with CCC and MEL; all coauthors reviewed the manuscript and approved its final version.

References

Abdul Wahid, A. and E. Arunbabu (2022) Forecasting water quality using seasonal ARIMA model by integrating in-situ measurements and remote sensing techniques in Krishnagiri reservoir, India. *Water Pract. Technol.*, **17**, 1230–52.

Baracchini, T., P. Y. Chu, J. Šukys, G. Lieberherr, S. Wunderle, A. Wüest, and D. Bouffard (2020a) Data assimilation of in situ and satellite remote sensing data to 3D hydrodynamic lake models: a case study using Delft3D-FLOW v4.03 and OpenDA v2.4. *Geosci. Model Dev.*, **13**, 1267–84.

Baracchini, T., A. Wüest, and D. Bouffard (2020b) Meteolakes: An operational online three-dimensional forecasting platform for lake hydrodynamics. *Water Res.*, **172**, 115529.

Bennett, N. D., B. F. Croke, G. Guariso, J. H. Guillaume, S. H. Hamilton, A. J. Jakeman, S. Marsili-Libelli, L. T. Newham, *et al.* (2013) Characterising performance of environmental models. *Environ. Model. Softw.*, **40**, 1–20.

Bett, P. E., K. E. Williams, C. Burton, A. A. Scaife, A. J. Wiltshire, and R. Gilham (2020)

Skillful seasonal prediction of key carbon cycle components: NPP and fire risk. *Environ. Res. Commun.*, **2**, 055002.

Bruce, L. C., M. A. Frassl, G. B. Arhonditsis, G. Gal, D. P. Hamilton, P. C. Hanson, A. L. Hetherington, J. M. Melack, *et al.* (2018) A multi-lake comparative analysis of the General Lake Model (GLM): Stress-testing across a global observatory network. *Environ. Model. Softw.*, **102**, 274–91.

Carey, C. C. and A. Breef-Pilz (2022) Ice cover data for Falling Creek Reservoir and Beaverdam Reservoir, Vinton, Virginia, USA for 2013-2022. *Environmental Data Initiative repository*. DOI: 10.6073/pasta/917b3947d91470eecf979e9297ed4d2e

Carey, C. C., A. Breef-Pilz, B. J. Bookout, R. P. McClure, and J. H. Wynne (2023) Time series of high-frequency sensor data measuring water temperature, dissolved oxygen, conductivity, specific conductance, total dissolved solids, chlorophyll a, phycocyanin, fluorescent dissolved organic matter, and turbidity at discrete depths in Beaverdam Reservoir, Virginia, USA in 2016-2022. *Environmental Data Initiative repository*. DOI: 10.6073/pasta/4182de376fde52e15d493fdd9f26d0c7

Carey, C. C., P. C. Hanson, R. Q. Thomas, A. B. Gerling, A. G. Hounshell, A. S. L. Lewis, M. E. Lofton, R. P. McClure, *et al.* (2022a) Anoxia decreases the magnitude of the carbon, nitrogen, and phosphorus sink in freshwaters. *Glob. Change Biol.*, **28**, 4861–81.

Carey, C. C., A. S. L. Lewis, D. W. Howard, W. M. Woelmer, P. A. Gantzer, K. A. Bierlein, J. C. Little, and WVWA (2022b) Bathymetry and watershed area for Falling Creek Reservoir, Beaverdam Reservoir, and Carvins Cove Reservoir. *Environmental Data Initiative repository*. DOI: 10.6073/pasta/352735344150f7e77d2bc18b69a22412

Carey, C. C., A. S. Lewis, R. P. McClure, A. B. Gerling, A. Breef-Pilz, and A. Das (2022c) Time series of high-frequency profiles of depth, temperature, dissolved oxygen, conductivity, specific conductance, chlorophyll a, turbidity, pH, oxidation-reduction potential, photosynthetic active radiation, and descent rate for Beaverdam Reservoir, Carvins Cove Reservoir, Falling Creek Reservoir, Gatewood Reservoir, and Spring Hollow Reservoir in Southwestern Virginia, USA 2013-2021. *Environmental Data Initiative repository*. DOI: 10.6073/pasta/c4c45b5b10b4cb4cd4b5e613c3effbd0

Carey, C. C., W. M. Woelmer, M. E. Lofton, R. J. Figueiredo, B. J. Bookout, R. S. Corrigan, V. Daneshmand, A. G. Hounshell, *et al.* (2022d) Advancing lake and reservoir water quality management with near-term, iterative ecological forecasting. *Inland Waters*, **12**, 107–20.

Carpenter, S. R., E. H. Stanley, and M. J. Vander Zanden (2011) State of the world's freshwater ecosystems: physical, chemical, and biological changes. *Annu. Rev. Environ. Resour.*, **36**, 75–99.

Chen, C., Q. Chen, G. Li, M. He, J. Dong, H. Yan, Z. Wang, and Z. Duan (2021) A novel multi-source data fusion method based on Bayesian inference for accurate estimation of chlorophyll-a concentration over eutrophic lakes. *Environ. Model. Softw.*, **141**, 105057.

Cho, H. and H. Park (2019) Merged-LSTM and multistep prediction of daily chlorophyll-a concentration for algal bloom forecast. *IOP Conf. Ser. Earth Environ. Sci.*, **351**, 012020.

Cho, K. H., Y. Pachepsky, M. Ligaray, Y. Kwon, and K. H. Kim (2020) Data assimilation in surface water quality modeling: A review. *Water Res.*, **186**, 116307.

Clark, J. S., S. R. Carpenter, M. Barber, S. Collins, A. Dobson, J. A. Foley, D. M. Lodge, M. Pascual, *et al.* (2001) Ecological forecasts: an emerging imperative. *science*, **293**, 657–60.

Clark, P., N. Roberts, H. Lean, S. P. Ballard, and C. Charlton-Perez (2016) Convectionpermitting models: a step-change in rainfall forecasting. *Meteorol. Appl.*, **23**, 165–81.

Clayer, F., L. Jackson-Blake, D. Mercado-Bettín, M. Shikhani, A. French, T. Moore, J. Sample, M. Norling, *et al.* (2023) Sources of skill in lake temperature, discharge and ice-off seasonal forecasting tools. *Hydrol. Earth Syst. Sci.*, **27**, 1361–81.

Corbari, C., R. Salerno, A. Ceppi, V. Telesca, and M. Mancini (2019) Smart irrigation forecast using satellite LANDSAT data and meteo-hydrological modeling. *Agric. Water Manag.*, **212**, 283–94.

Daneshmand, V., A. Breef-Pilz, C. C. Carey, Y. Jin, Y.-J. Ku, K. C. Subratie, R. Q. Thomas, and R. J. Figueiredo (2021) Edge-to-cloud Virtualized Cyberinfrastructure for Near Real-time Water Quality Forecasting in Lakes and Reservoirs. *2021 IEEE 17th International Conference on eScience (eScience)*. IEEE, pp. 138–48.

DeChant, C. M. and H. Moradkhani (2011) Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation. *Hydrol. Earth Syst. Sci.*, **15**, 3399–3410.

Derot, J., H. Yajima, and F. G. Schmitt (2020) Benefits of machine learning and sampling frequency on phytoplankton bloom forecasts in coastal areas. *Ecol. Inform.*, **60**, 101174.

Dietze, M. C. (2017a) *Ecological forecasting*. Princeton University Press.

Dietze, M. C. (2017b) Prediction in ecology: a first-principles framework. *Ecol. Appl.*, **27**, 2048–60.

Dietze, M. C., A. Fox, L. M. Beck-Johnson, J. L. Betancourt, M. B. Hooten, C. S. Jarnevich, T. H. Keitt, M. A. Kenney, *et al.* (2018) Iterative near-term ecological forecasting: Needs, opportunities, and challenges. *Proc. Natl. Acad. Sci.*, **115**, 1424–32.

Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins, C. D.

Bertelsen, and D. W. Inouye (2012) Forecasting phenology: from species variability to community patterns. *Ecol. Lett.*, **15**, 545–53.

Doubek, J. P., K. L. Campbell, M. E. Lofton, R. P. McClure, and C. C. Carey (2019) Hypolimnetic hypoxia increases the biomass variability and compositional variability of crustacean zooplankton communities. *Water*, **11**, 2179.

Duc, L., K. Saito, and D. Hotta (2021) Analysis and design of covariance inflation methods using inflation functions. Part 2: adaptive inflation. *Q. J. R. Meteorol. Soc.*, **147**, 2375–94.

Engelhardt, C. and G. Kirillin (2014) Criteria for the onset and breakup of summer lake stratification based on routine temperature measurements. *Fundam. Appl. Limnol.*, **184**, 183–94.

Evensen, G. (2003) The Ensemble Kalman Filter: theoretical formulation and practical implementation. *Ocean Dyn.*, **53**, 343–67.

Francy, D. S., J. L. Graham, E. A. Stelzer, C. D. Ecker, A. M. Brady, P. Struffolino, and K. A. Loftin (2015) Water quality, cyanobacteria, and environmental factors and their relations to microcystin concentrations for use in predictive models at ohio lake erie and inland lake recreational sites, 2013-14. US Geological Survey Scientific Investigations Report 2015–5120, 58 p.

Georgakakos, K. P., N. E. Graham, T. M. Carpenter, and H. Yao (2005) Integrating climate-hydrology forecasts and multi-objective reservoir management for northern California. *Eos Trans. Am. Geophys. Union*, **86**, 122–27.

George, D. G. and M. A. Hurley (2004) The influence of sampling frequency on the detection of long-term change in three lakes in the English Lake District. *Aquat. Ecosyst. Health*

Manag., **7**, 1–14.

Gilarranz, L. J., A. Narwani, D. Odermatt, R. Siber, and V. Dakos (2022) Regime shifts, trends, and variability of lake productivity at a global scale. *Proc. Natl. Acad. Sci.*, **119**, e2116413119.

Gneiting, T., A. E. Raftery, A. H. Westveld, and T. Goldman (2005) Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Mon. Weather Rev.*, **133**, 1098–1118.

Gottwald, G. A. and S. Reich (2021) Supervised learning from noisy observations: Combining machine-learning techniques with data assimilation. *Phys. Nonlinear Phenom.*, **423**, 132911.

Grönquist, P., C. Yao, T. Ben-Nun, N. Dryden, P. Dueben, S. Li, and T. Hoefler (2021) Deep learning for post-processing ensemble weather forecasts. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.*, **379**, 20200092.

Hamre, K. D., M. E. Lofton, R. P. McClure, Z. W. Munger, J. P. Doubek, A. B. Gerling, M. E. Schreiber, and C. C. Carey (2018) In situ fluorometry reveals a persistent, perennial hypolimnetic cyanobacterial bloom in a seasonally anoxic reservoir. *Freshw. Sci.*, **37**, 483–95.

Harris, D. J., S. D. Taylor, and E. P. White (2018) Forecasting biodiversity in breeding birds using best practices. *PeerJ*, **6**, e4278.

He, H., L. Lei, J. S. Whitaker, and Z.-M. Tan (2020) Impacts of Assimilation Frequency on Ensemble Kalman Filter Data Assimilation and Imbalances. *J. Adv. Model. Earth Syst.*, **12**, e2020MS002187.

Heilman, K. A., M. C. Dietze, A. A. Arizpe, J. Aragon, A. Gray, J. D. Shaw, A. O. Finley, S.

Klesse, *et al.* (2022) Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty. *Glob. Change Biol.*, **28**, 2442–60.

Herrick, C., B. G. Steele, J. A. Brentrup, K. L. Cottingham, M. J. Ducey, D. A. Lutz, M. W. Palace, M. C. Thompson, *et al.* (2023) lakeCoSTR: A tool to facilitate use of Landsat Collection 2 to estimate lake surface water temperatures. *Ecosphere*, **14**, e4357.

Hipsey, M. R., C. Boon, L. C. Bruce, Q. Thomas, M. Weber, L. Winslow, J. S. Read, and D. P. Hamilton (2022) AquaticEcoDynamics/glm-aed: v3.3.0.

Hipsey, M. R., L. C. Bruce, C. Boon, B. Busch, C. C. Carey, D. P. Hamilton, P. C. Hanson, J. S. Read, *et al.* (2019) A General Lake Model (GLM 3.0) for linking with high-frequency sensor data from the Global Lake Ecological Observatory Network (GLEON). *Geosci. Model Dev.*, **12**, 473–523.

Hounshell, A. G., R. P. McClure, M. E. Lofton, and C. C. Carey (2021) Whole-ecosystem oxygenation experiments reveal substantially greater hypolimnetic methane concentrations in reservoirs during anoxia. *Limnol. Oceanogr. Lett.*, **6**, 33–42.

Ibáñez, I., E. S. Gornish, L. Buckley, D. M. Debinski, J. Hellmann, B. Helmuth, J. HilleRisLambers, A. M. Latimer, *et al.* (2013) Moving forward in global-change ecology: capitalizing on natural variability. *Ecol. Evol.*, **3**, 170–81.

Jolliffe, I. T. and D. B. Stephenson (2012) *Forecast verification: a practitioner's guide in atmospheric science*. John Wiley & Sons.

Kehoe, M. J., K. P. Chun, and H. M. Baulch (2015) Who Smells? Forecasting Taste and Odor in a Drinking Water Reservoir. *Environ. Sci. Technol.*, **49**, 10984–92.

Kirchner, J. W. and C. Neal (2013) Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. *Proc. Natl. Acad. Sci.*, **110**, 12213–18.

LaDeau, S. L., B. A. Han, E. J. Rosi-Marshall, and K. C. Weathers (2017) The Next Decade of Big Data in Ecosystem Science. *Ecosystems*, **20**, 274–83.

Ladwig, R., L. A. Rock, and H. A. Dugan (2021) Impact of salinization on lake stratification and spring mixing. *Limnol. Oceanogr. Lett, **8**, 93-102.*

Lewis, A. S. L., W. M. Woelmer, H. L. Wander, D. W. Howard, J. W. Smith, R. P. McClure, M. E. Lofton, N. W. Hammond, *et al.* (2022) Increased adoption of best practices in ecological forecasting enables comparisons of forecastability. *Ecol. Appl.*, **32**, e2500.

Lin, E., Y. Yang, X. Qiu, Q. Xie, R. Gan, B. Zhang, and X. Liu (2021) Impacts of the radar data assimilation frequency and large-scale constraint on the short-term precipitation forecast of a severe convection case. *Atmospheric Res.*, **257**, 105590.

Lindegren, M., C. Möllmann, A. Nielsen, K. Brander, B. R. MacKenzie, and N. Chr. Stenseth (2010) Ecological forecasting under climate change: the case of Baltic cod. *Proc. R. Soc. B Biol. Sci.*, **277**, 2121–30.

Liu, H., Y. D. Chen, T. Liu, and L. Lin (2019) The river chief system and river pollution control in China: A case study of Foshan. *Water*, **11**, 1606.

Lofton, M. E., J. A. Brentrup, W. S. Beck, J. A. Zwart, R. Bhattacharya, L. S. Brighenti, S. H. Burnet, I. M. McCullough, *et al.* (2022) Using near-term forecasts and uncertainty partitioning to inform prediction of oligotrophic lake cyanobacterial density. *Ecol. Appl.*, **32**, e2590.

Lofton, M. E., D. W. Howard, R. Q. Thomas, and C. C. Carey (2023) Progress and opportunities in advancing near-term forecasting of freshwater quality. *Glob. Change Biol.*, **29**,1691–

1714.
Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel (2011) Ecological forecasting and data assimilation in a data-rich era. *Ecol. Appl.*, **21**, 1429–42.
Machete, R. L. and L. A. Smith (2016) Demonstrating the value of larger ensembles in forecasting physical systems. *Tellus Dyn. Meteorol. Oceanogr.*, **68**, 28393.
Magnuson, J. J., L. B. Crowder, and P. A. Medvick (1979) Temperature as an ecological resource. *Am. Zool.*, **19**, 331–43.
Malhi, Y., J. Franklin, N. Seddon, M. Solan, M. G. Turner, C. B. Field, and N. Knowlton (2020) Climate change and ecosystems: threats, opportunities and solutions. *Philos. Trans. R. Soc. B Biol. Sci.*, **375**, 20190104.
Mantovani, C., L. Corgnati, J. Horstmann, A. Rubio, E. Reyes, C. Quentin, S. Cosoli, J. L. Asensio, *et al.* (2020) Best Practices on High Frequency Radar Deployment and Operation for Ocean Current Measurement. *Front. Mar. Sci.*, **7,** 210.
Marcé, R., G. George, P. Buscarinu, M. Deidda, J. Dunalska, E. de Eyto, G. Flaim, H.-P. Grossart, *et al.* (2016) Automatic High Frequency Monitoring for Improved Lake and Reservoir Management. *Environ. Sci. Technol.*, **50**, 10780–94.
Marj, A. F. and A. M. J. Meijerink (2011) Agricultural drought forecasting using satellite images, climate indices and artificial neural network. *Int. J. Remote Sens.*, **32**, 9707–19.
Massoud, E. C., J. Huisman, E. Benincà, M. C. Dietze, W. Bouten, and J. A. Vrugt (2018) Probing the limits of predictability: data assimilation of chaotic dynamics in complex food webs. *Ecol. Lett.*, **21**, 93–103.
McClure, R. P., R. Q. Thomas, M. E. Lofton, W. M. Woelmer, and C. C. Carey (2021) Iterative Forecasting Improves Near-Term Predictions of Methane Ebullition Rates. *Front.*

Environ. Sci., **9**, 756603.

Mercado-Bettín, D., F. Clayer, M. Shikhani, T. N. Moore, M. D. Frías, L. Jackson-Blake, J. Sample, M. Iturbide, *et al.* (2021) Forecasting water temperature in lakes and reservoirs using seasonal climate prediction. *Water Res.*, **201**, 117286.

Meyer, J. L., M. J. Sale, P. J. Mulholland, and N. L. Poff (1999) Impacts of climate change on aquatic ecosystem functioning and health 1. *JAWRA J. Am. Water Resour. Assoc.*, **35**, 1373–86.

Mi, C., T. Shatwell, J. Ma, Y. Xu, F. Su, and K. Rinke (2020) Ensemble warming projections in Germany's largest drinking water reservoir and potential adaptation strategies. *Sci. Total Environ.*, **748**, 141366.

Moustahfid, H., L. C. Hendrickson, A. Arkhipkin, G. J. Pierce, A. Gangopadhyay, H. Kidokoro, U. Markaida, C. Nigmatullin, *et al.* (2021) Ecological-fishery forecasting of squid stock dynamics under climate variability and change: review, Challenges, and Recommendations. *Rev. Fish. Sci. Aquac.*, **29**, 682–705.

Niu, S., Y. Luo, M. C. Dietze, T. F. Keenan, Z. Shi, J. Li, and F. S. C. Iii (2014) The role of data assimilation in predictive ecology. *Ecosphere*, **5**, 1-16.

O'Reilly, C. M., S. Sharma, D. K. Gray, S. E. Hampton, J. S. Read, R. J. Rowley, P. Schneider, J. D. Lenters, *et al.* (2015) Rapid and highly variable warming of lake surface waters around the globe. *Geophys. Res. Lett.*, **42**, 10–773.

Paerl, H. W. and V. J. Paul (2012) Climate change: links to global expansion of harmful cyanobacteria. *Water Res.*, **46**, 1349–63.

Page, T., P. J. Smith, K. J. Beven, I. D. Jones, J. A. Elliott, S. C. Maberly, E. B. Mackay, M. De Ville, *et al.* (2018) Adaptive forecasting of phytoplankton communities. *Water Res.*, **134**, 74–85.

Page, T., P. J. Smith, K. J. Beven, I. D. Jones, J. A. Elliott, S. C. Maberly, E. B. Mackay, M. De Ville, *et al.* (2017) Constraining uncertainty and process-representation in an algal community lake model using high frequency in-lake observations. *Ecol. Model.*, **357**, 1–13.

Park, J., K. T. Kim, and W. H. Lee (2020) Recent Advances in Information and Communications Technology (ICT) and Sensor Technology for Monitoring Water Quality. *Water*, **12**, 510.

Parrish, M. A., H. Moradkhani, and C. M. DeChant (2012) Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation. *Water Resour. Res.*, **48,** W03519.

Piazzi, G., G. Thirel, L. Campo, and S. Gabellani (2018) A particle filter scheme for multivariate data assimilation into a point-scale snowpack model in an Alpine environment. *The Cryosphere*, **12**, 2287–2306.

R Core Team (2022) R: A language and environment for statistical computing.

Read, J. S., X. Jia, J. Willard, A. P. Appling, J. A. Zwart, S. K. Oliver, A. Karpatne, G. J. A. Hansen, *et al.* (2019) Process-Guided Deep Learning Predictions of Lake Water Temperature. *Water Resour. Res.*, **55**, 9173–90.

Read, J. S., L. A. Winslow, G. J. A. Hansen, J. Van Den Hoek, P. C. Hanson, L. C. Bruce, and C. D. Markfort (2014) Simulating 2368 temperate lakes reveals weak coherence in stratification phenology. *Ecol. Model.*, **291**, 142–50.

Romero, J. R., I. Kagalou, J. Imberger, D. Hela, M. Kotti, A. Bartzokas, T. Albanis, N. Evmirides, *et al.* (2002) Seasonal water quality of shallow and eutrophic Lake Pamvotis, Greece: implications for restoration. *Hydrobiologia*, **474**, 91-105

Rousso, B. Z., E. Bertone, R. Stewart, and D. P. Hamilton (2020) A systematic literature review of forecasting and predictive models for cyanobacteria blooms in freshwater lakes. *Water Res.*, **182**, 115959.

Simonin, D., C. Pierce, N. Roberts, S. P. Ballard, and Z. Li (2017) Performance of Met Office hourly cycling NWP-based nowcasting for precipitation forecasts. *Q. J. R. Meteorol. Soc.*, **143**, 2862–73.

Steere, D. C., A. Baptista, D. McNamee, C. Pu, and J. Walpole (2000) Research challenges in environmental observation and forecasting systems. *Proceedings of the 6th annual international conference on Mobile computing and networking - MobiCom '00*. ACM Press, Boston, Massachusetts, United States, pp. 292–99.

Tanut, B., R. Waranusast, and P. Riyamongkol (2021) High accuracy pre-harvest sugarcane yield forecasting model utilizing drone image analysis, data mining, and reverse design method. *Agriculture*, **11**, 682.

Thomas, R. Q., C. Boettiger, C. C. Carey, M. C. Dietze, L. R. Johnson, M. A. Kenney, J. S. McLachlan, J. A. Peters, *et al.* (2023a) The NEON Ecological Forecasting Challenge. *Front. Ecol. Environ.*, **21**, 112–13.

Thomas, R. Q., R. J. Figueiredo, V. Daneshmand, B. J. Bookout, L. K. Puckett, and C. C. Carey (2020) A Near-Term Iterative Forecasting System Successfully Predicts Reservoir Hydrodynamics and Partitions Uncertainty in Real Time. *Water Resour. Res.*, **56**, e2019WR026138.

Thomas, R. Q., R. P. McClure, T. N. Moore, W. M. Woelmer, C. Boettiger, R. J. Figueiredo, R. T. Hensley, and C. C. Carey (2023b) Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the US. *Front. Ecol. Environ.*, **n/a**.

Wander, H. L., R. Q. Thomas, T. N. Moore, M. E. Lofton, A. Breef-Pilz, and C. C. Carey (2023a) Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir: data, forecasts, and scores [Data set]. *Zenodo*. https://doi.org/10.5281/zenodo.7951402

Wander, H. L., R. Q. Thomas, T. N. Moore, M. E. Lofton, A. Breef-Pilz, and C. C. Carey (2023b) hlwander/BVRE-forecast-code: Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir: Code (v1.1). *Zenodo*. https://doi.org/10.5281/zenodo.7958471

Wang, S., N. Flipo, and T. Romary (2023) Which filter for data assimilation in water quality models? Focus on oxygen reaeration and heterotrophic bacteria activity. *J. Hydrol.*, **620**, 129423.

Wang, X., J. Zhang, and V. Babovic (2016) Improving real-time forecasting of water quality indicators with combination of process-based models and data assimilation technique. *Ecol. Indic.*, **66**, 428–39.

Weng, E. and Y. Luo (2011) Relative information contributions of model vs. data to short- and long-term forecasts of forest carbon dynamics. *Ecol. Appl.*, **21**, 1490–1505.

Wheeler, K., M. Dietze, D. LeBauer, J. Peters, A. D. Richardson, R. Q. Thomas, K. Zhu, U. Bhat, *et al.* (2023) Predicting Spring Phenology in Deciduous Broadleaf Forests: An Open Community Forecast Challenge. Available at SSRN: 10.2139/ssrn.4357147

White, E. P., G. M. Yenni, S. D. Taylor, E. M. Christensen, E. K. Bledsoe, J. L. Simonis, and S. K. M. Ernest (2019) Developing an automated iterative near-term forecasting system for an ecological study. *Methods Ecol. Evol.*, **10**, 332–44.

Williamson, C. E., E. P. Overholt, J. A. Brentrup, R. M. Pilla, T. H. Leach, S. G. Schladow, J. D.

Warren, S. S. Urmy, *et al.* (2016) Sentinel responses to droughts, wildfires, and floods: effects of UV radiation on lakes and their ecosystem services. *Front. Ecol. Environ.*, **14**, 102–9.

Woelmer, W. M., R. Q. Thomas, M. E. Lofton, R. P. McClure, H. L. Wander, and C. C. Carey (2022) Near-term phytoplankton forecasts reveal the effects of model time step and forecast horizon on predictability. *Ecol. Appl.*, **32**, e2642.

Woolway, R. I., E. Jennings, T. Shatwell, M. Golub, D. C. Pierson, and S. C. Maberly (2021) Lake heatwaves under climate change. *Nature*, **589**, 402–7.

Yvon-Durocher, G., J. M. Caffrey, A. Cescatti, M. Dossena, P. del Giorgio, J. M. Gasol, J. M. Montoya, J. Pumpanen, *et al.* (2012) Reconciling the temperature dependence of respiration across timescales and ecosystem types. *Nature*, **487**, 472–76.

Ziliani, M. G., R. Ghostine, B. Ait-El-Fquih, M. F. McCabe, and I. Hoteit (2019) Enhanced flood forecasting through ensemble data assimilation and joint state-parameter estimation. *J. Hydrol.*, **577**, 123924.

Figure Captions

Figure 1: Forecasting Lake And Reservoir Ecosystems (FLARE) workflow showing the step-by-step process for generating daily water temperature forecasts, starting with data collection from thermistors deployed in the reservoir (step 1), then data access for running the forecast model (step 2), then generation of forecasts with data assimilation (step 3), and ending with forecast skill assessment (step 4). During the data assimilation steps (3a-b), data assimilation experiments were performed with four different data assimilation frequencies (daily, weekly, fortnightly, and monthly; see dashed line box). Steps 1-4 occurred throughout the entire forecast period (1 January - 31 December 2021). Buoy figure via NexSens Technology Inc., CC by 2.0 https://creativecommons.org/licenses/by/2.0/

Figure 2: Map of Beaverdam Reservoir, Vinton, VA (37.31° N, 79.82° W). The map shows the surrounding forested watershed; the point represents the reservoir monitoring site where high-frequency sensor data were collected.

Figure 3: Observed water temperature for all depths with high-frequency sensors during the forecasting period of 1 January - 31 December 2021 in Beaverdam Reservoir (BVR). The gray background indicates the mixed period (1 January - 11 March, 8 November - 31 December 2021), while the white background indicates the thermally-stratified period (12 March - 7 November 2021), defined by a <0.1 kg/m^3 density differential between surface and bottom layers.

Figure 4: Example of water temperature forecasts at 1 m (a), 5 m (b), and 9 m (c) generated for 1-35 days into the future in Beaverdam Reservoir. Data assimilation (DA) frequencies are depicted by colors; shading shows 95% confidence intervals around the mean predicted temperature for each day. Black points represent water temperature observations. Colored points represent the most recent day that data was assimilated for each DA frequency. In this example, data were most recently assimilated on the day that the forecasts were generated: 25 June for the monthly DA scenario, 9 July for the fortnightly DA scenario, 16 July for the weekly DA scenario, and 22 July for the daily DA scenario.

Figure 5: Parameter evolution during the forecast period (1 January - 31 December 2021) for daily, weekly, fortnightly, and monthly data assimilation (DA) frequencies at 1-day-ahead forecast horizons. Longwave (a) is the longwave radiation scaling parameter, hypo_sed_temp (b) is the hypolimnetic sediment temperature parameter, and epi_sed_temp (c) is the epilimnetic sediment temperature parameter.

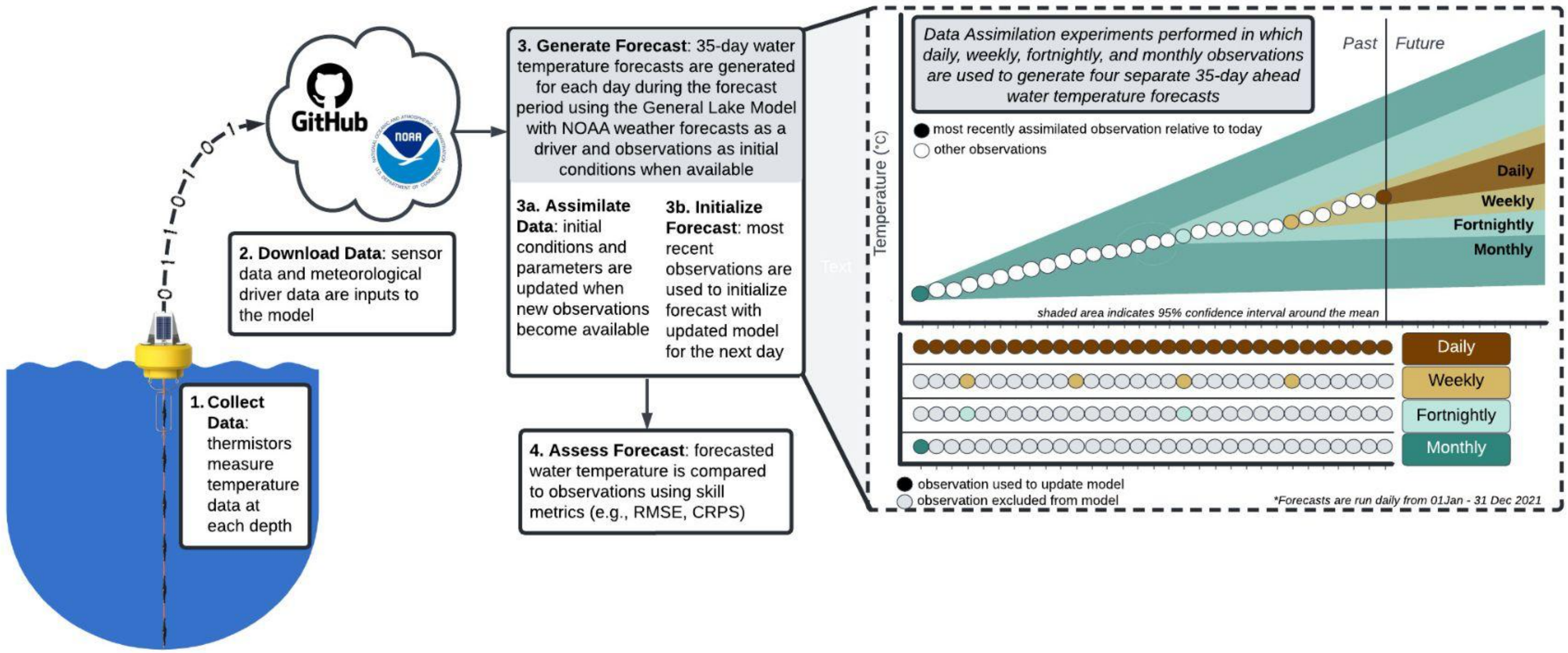
Figure 6: Root mean square error (RMSE) of mean forecasted water temperature compared to observations for 1-35 day-ahead forecast horizons in Beaverdam Reservoir, aggregated for all depths in the water column and days within the 365-day forecast period. RMSE for each forecast horizon was averaged from forecasts generated during 1 January - 31 December 2021. Colored lines represent different data assimilation (DA) frequencies. The dotted line depicts the 2°C threshold for skillful water temperature forecasts.

Figure 7: Root mean square error (RMSE) of mean forecasted water temperature compared to observations for 1-35-day-ahead forecast horizons in Beaverdam Reservoir during the mixed (panels a, c, e) vs. stratified (panels b, d, f) periods at 1 m (a, b), 5 m (c, d), and 9 m (e, f). RMSE for each forecast horizon was averaged across the 365-day forecast period (1 January - 31 December 2021). Colored lines correspond to different data assimilation (DA) frequencies (daily, weekly, fortnightly, and monthly); dotted horizontal lines depict the 2°C threshold for skillful forecasts.

Figure 8: Mean water temperature forecast variance across horizons (1-35 days ahead) in Beaverdam Reservoir during the mixed (panels a, c, e) vs. stratified (panels b, d, f) periods for 1 m (a, b), 5 m (c, d), and 9 m (e, f). Variance for each forecast horizon was averaged from all 365 forecasts generated during the forecast period (1 January - 31 December 2021). Colored lines correspond to different data assimilation (DA) frequencies (daily, weekly, fortnightly, and monthly).

Figure 9: Proportion of initial conditions uncertainty relative to total forecast uncertainty averaged across all forecasts generated with each DA frequency calculated using forecast variance across 1 January - 31 December 2021. Colored lines depict data assimilation (DA) frequencies (daily, weekly, fortnightly, and monthly). Panels a, c, and e represent mixed period forecasts, panels b, d, and f represent stratified period forecasts. Depths (1, 5, and 9 m) are indicated by gray facet labels.

Figure 1.



0–1–1–0–1–0–1–0–1
GitHub
NOAA
1. Collect Data: thermistors measure temperature data at each depth
2. Download Data: sensor data and meteorological driver data are inputs to the model
3. Generate Forecast: 35-day water temperature forecasts are generated for each day during the forecast period using the General Lake Model with NOAA weather forecasts as a driver and observations as initial conditions when available
3a. Assimilate Data: initial conditions and parameters are updated when new observations become available
3b. Initialize Forecast: most recent observations are used to initialize forecast with updated model for the next day
4. Assess Forecast: forecasted water temperature is compared to observations using skill metrics (e.g., RMSE, CRPS)
Data Assimilation experiments performed in which daily, weekly, fortnightly, and monthly observations are used to generate four separate 35-day ahead water temperature forecasts
Past
Future
most recently assimilated observation relative to today
other observations
Temperature (°C)
Daily
Weekly
Fortnightly
Monthly
shaded area indicates 95% confidence interval around the mean
Daily
Weekly
Fortnightly
Monthly
observation used to update model
observation excluded from model
*Forecasts are run daily from 01Jan - 31 Dec 2021

Figure 2.

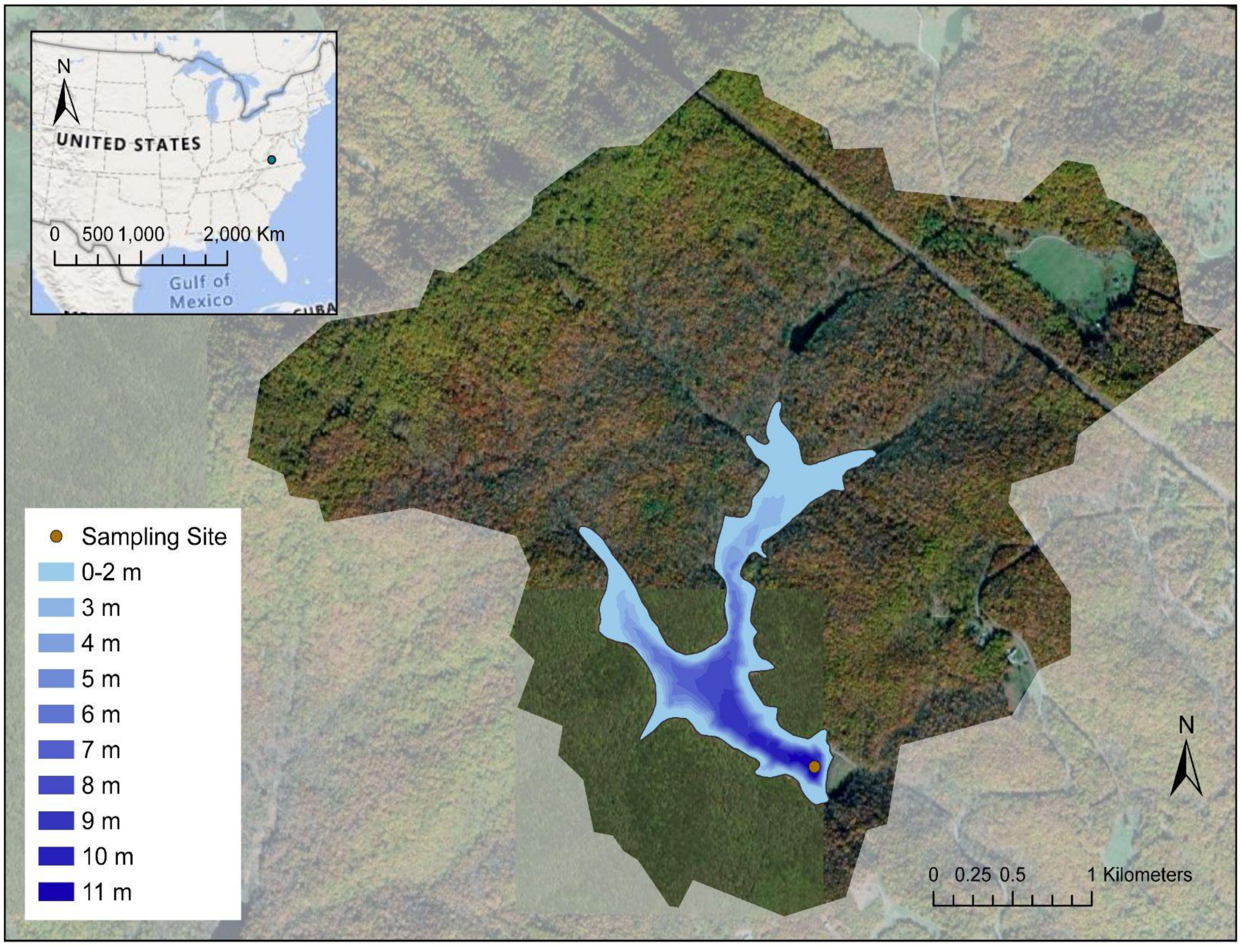
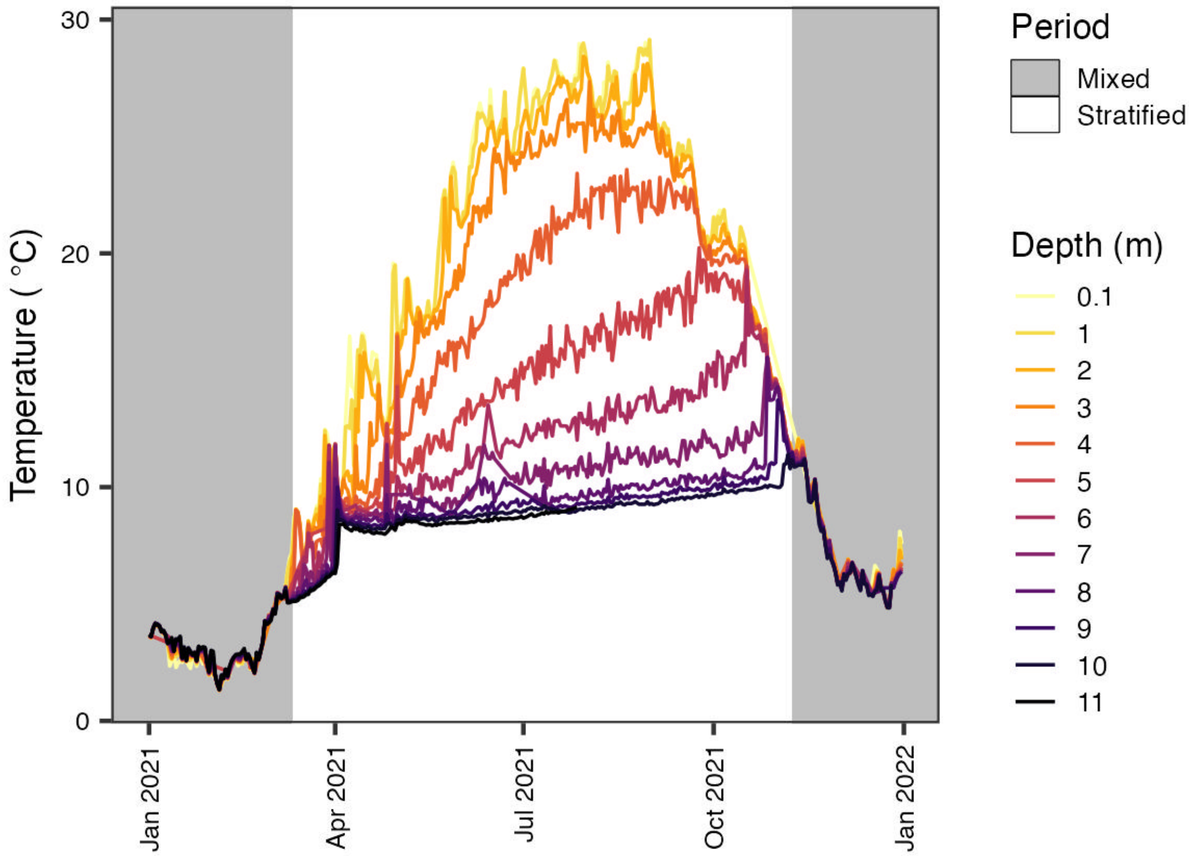
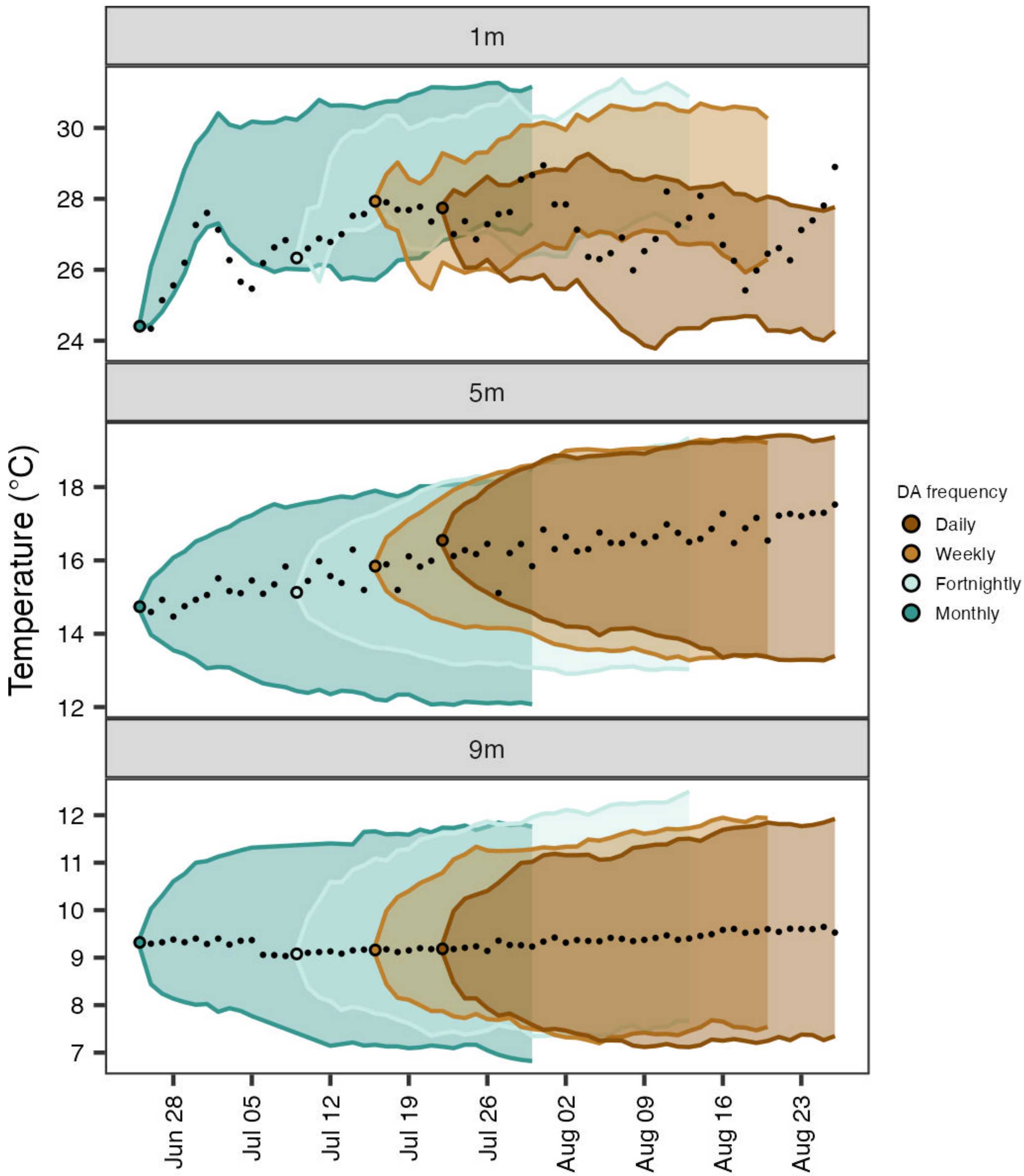

N
UNITED STATES
0 500 1,000 2,000 Km
Gulf of Mexico
CUBA
Sampling Site
0-2 m
3 m
4 m
5 m
6 m
7 m
8 m
9 m
10 m
11 m
N
0 0.25 0.5 1 Kilometers

Figure 3.



Temperature (°C)
30
20
10
0
Jan 2021
Apr 2021
Jul 2021
Oct 2021
Jan 2022
Period
Mixed
Stratified
Depth (m)
0.1
1
2
3
4
5
6
7
8
9
10
11

1076 Figure 4.



1077

Figure 5.

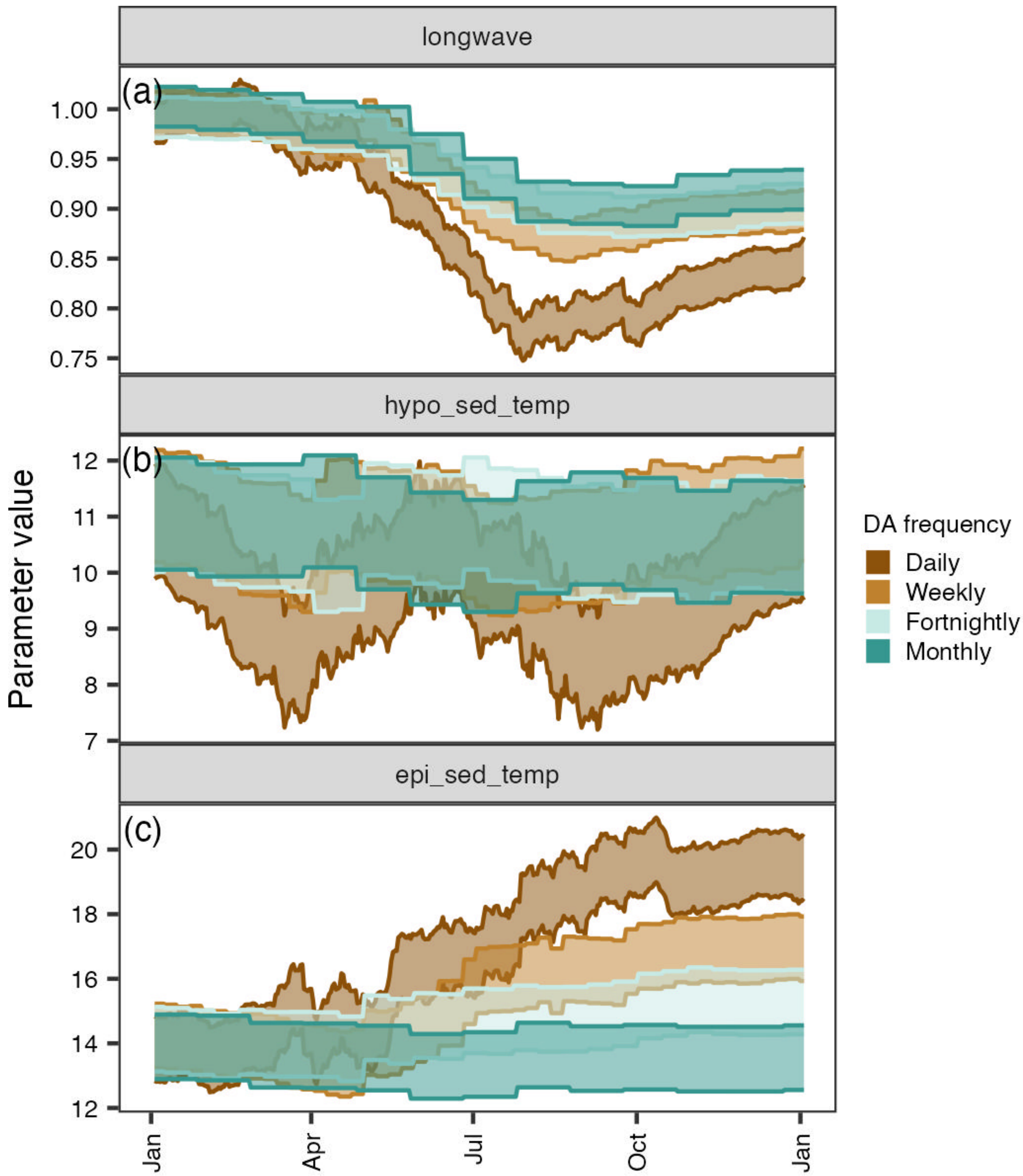
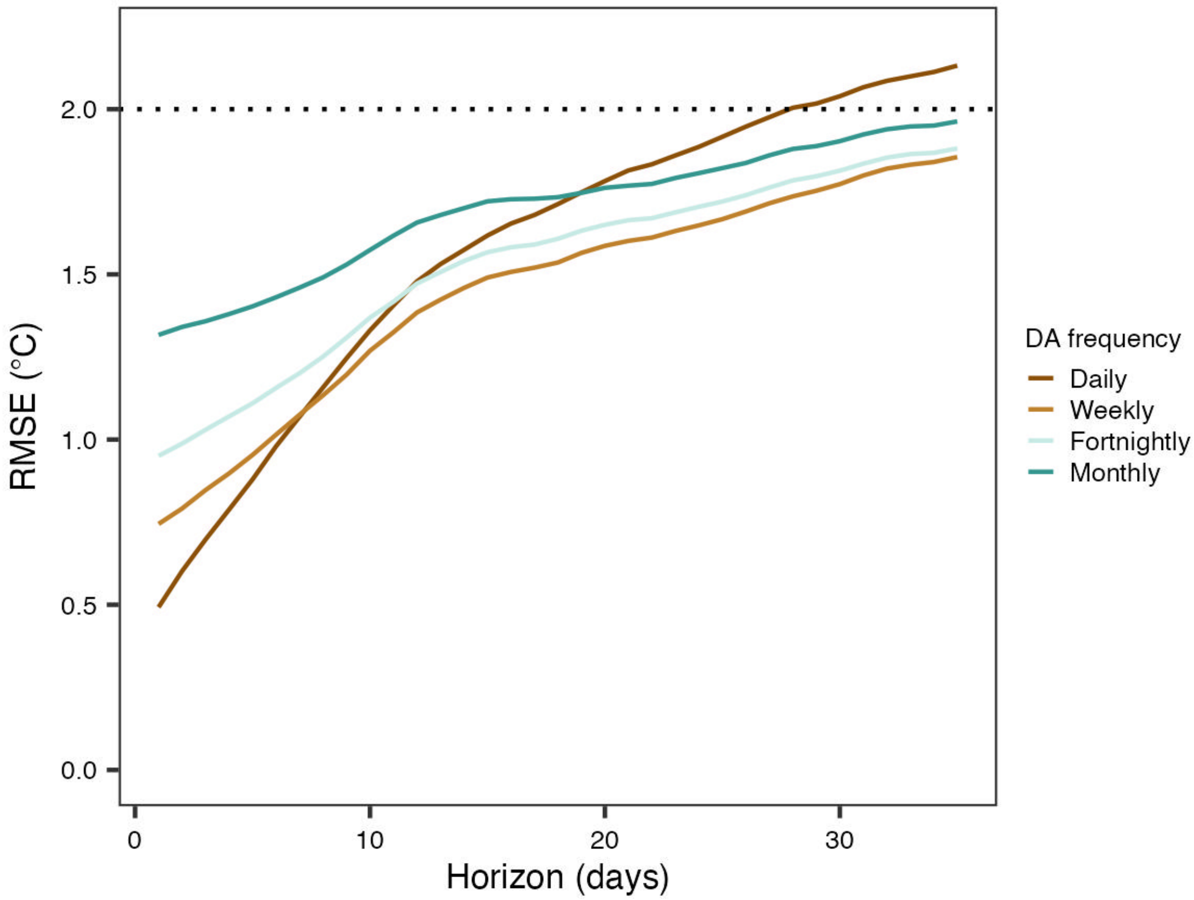
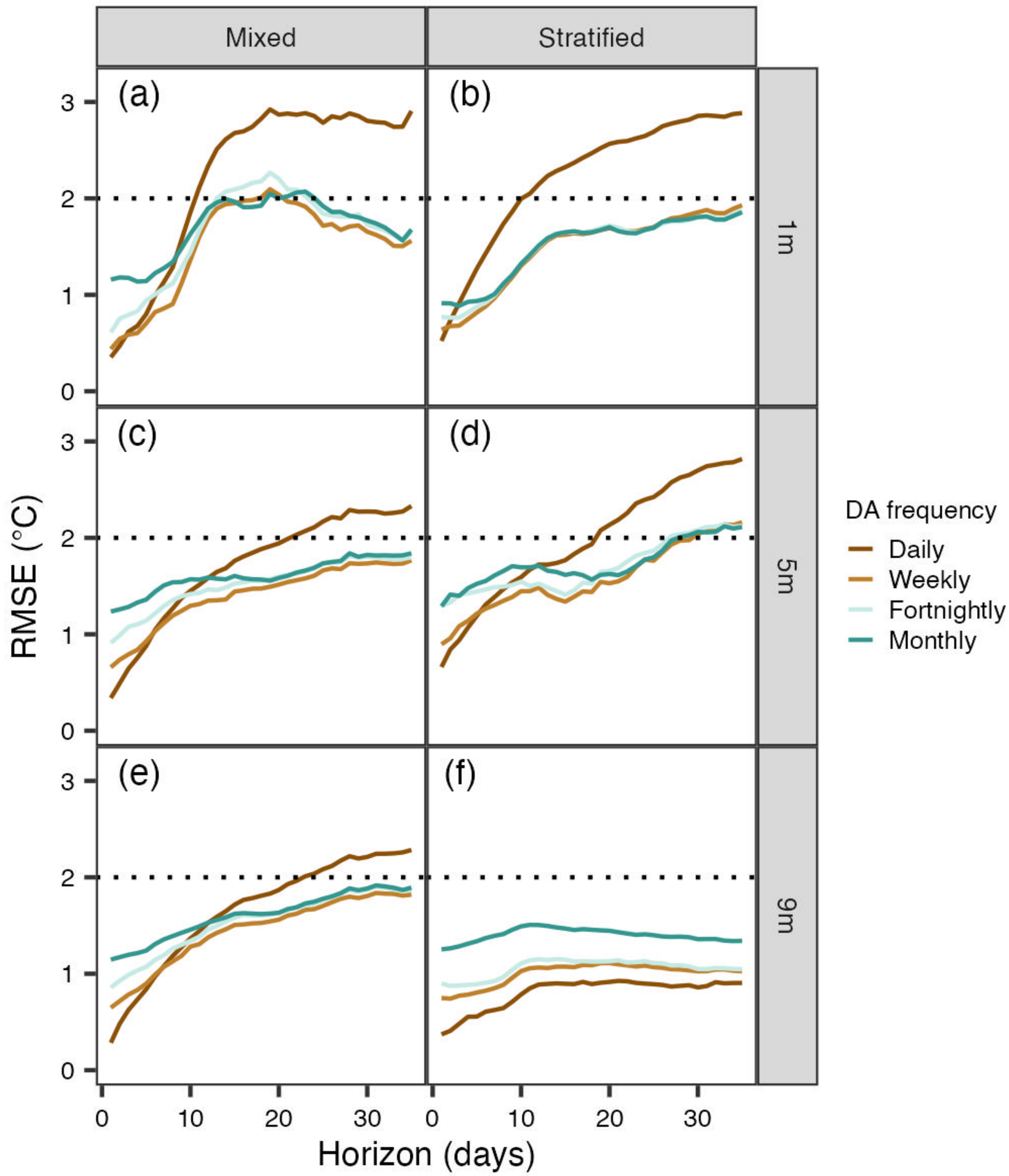

longwave
(a)
1.00
0.95
0.90
0.85
0.80
0.75
hypo_sed_temp
(b)
12
11
10
9
8
7
epi_sed_temp
(c)
20
18
16
14
12
Parameter value
Jan
Apr
Jul
Oct
Jan
DA frequency
Daily
Weekly
Fortnightly
Monthly

Figure 6.



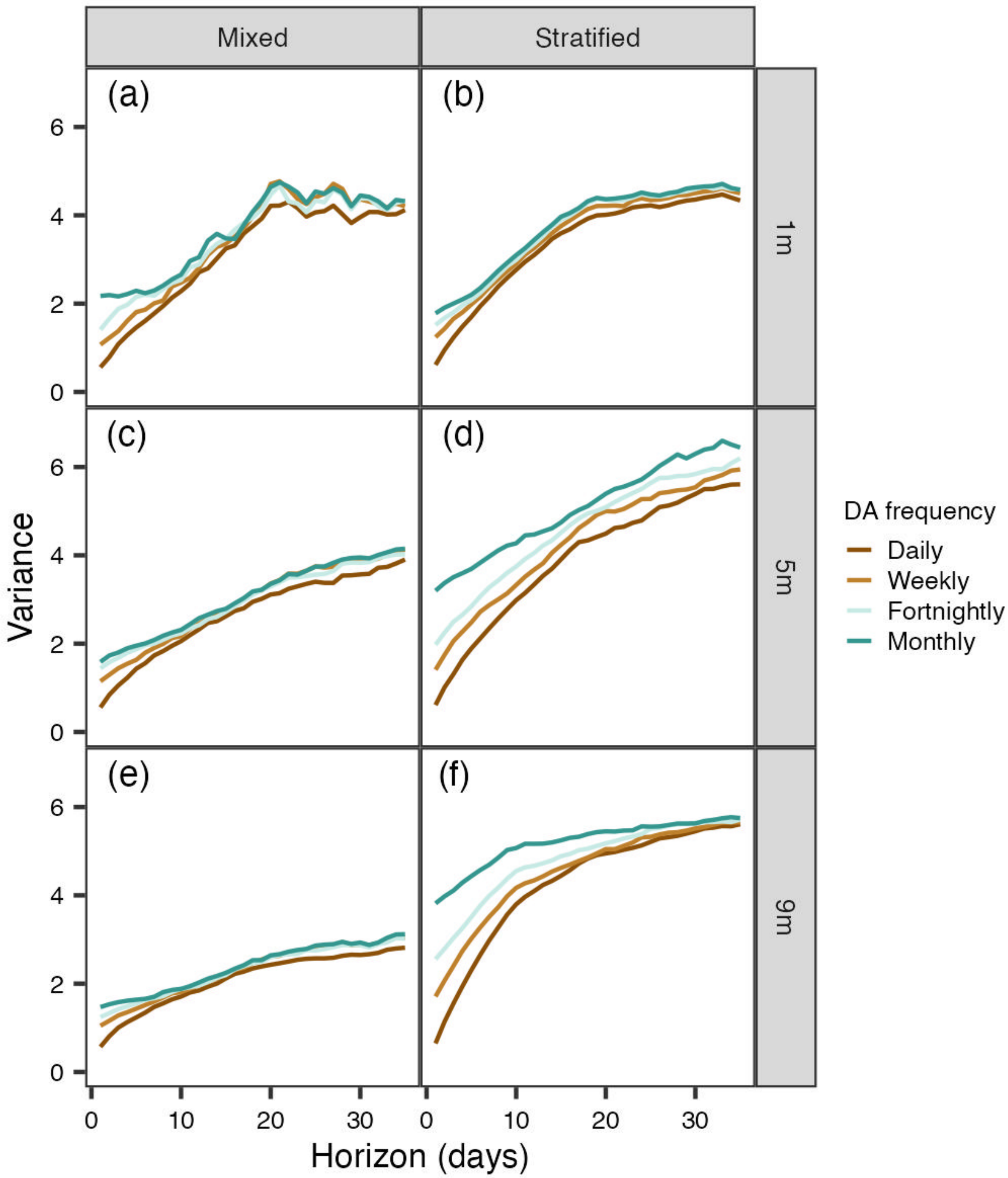
RMSE (°C)
2.0
1.5
1.0
0.5
0.0
0
10
20
30
Horizon (days)
DA frequency
Daily
Weekly
Fortnightly
Monthly

1082 Figure 7.



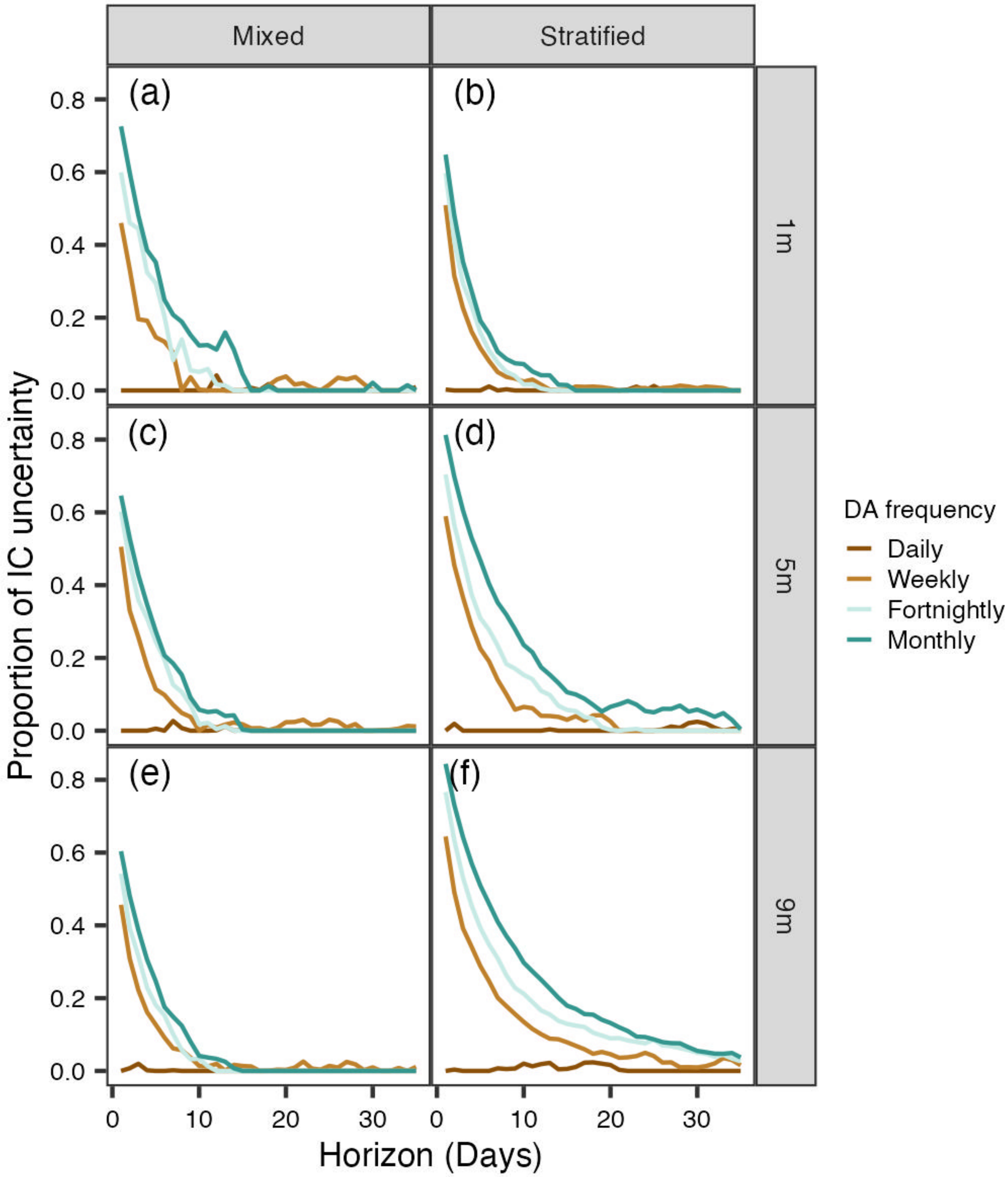
1083

Figure 8.



Mixed
Stratified
(a)
(b)
(c)
(d)
(e)
(f)
1m
5m
9m
Variance
Horizon (days)
DA frequency
Daily
Weekly
Fortnightly
Monthly

1086 Figure 9.



1087

Supplement to: **Data assimilation experiments inform monitoring needs for near-term ecological forecasts in a eutrophic reservoir**

Heather L. Wander[1], R. Quinn Thomas[1,2], Tadhg N. Moore[1,2], Mary E. Lofton[1], Adrienne Breef-Pilz[1], Cayelan C. Carey[1]

[1]Department of Biological Sciences, Virginia Tech, Blacksburg, VA

[2]Department of Forest Resources and Environmental Conservation, Virginia Tech, Blacksburg, VA

Submitted to *Ecosphere*: Methods, Tools, and Technologies Track

Appendix S1

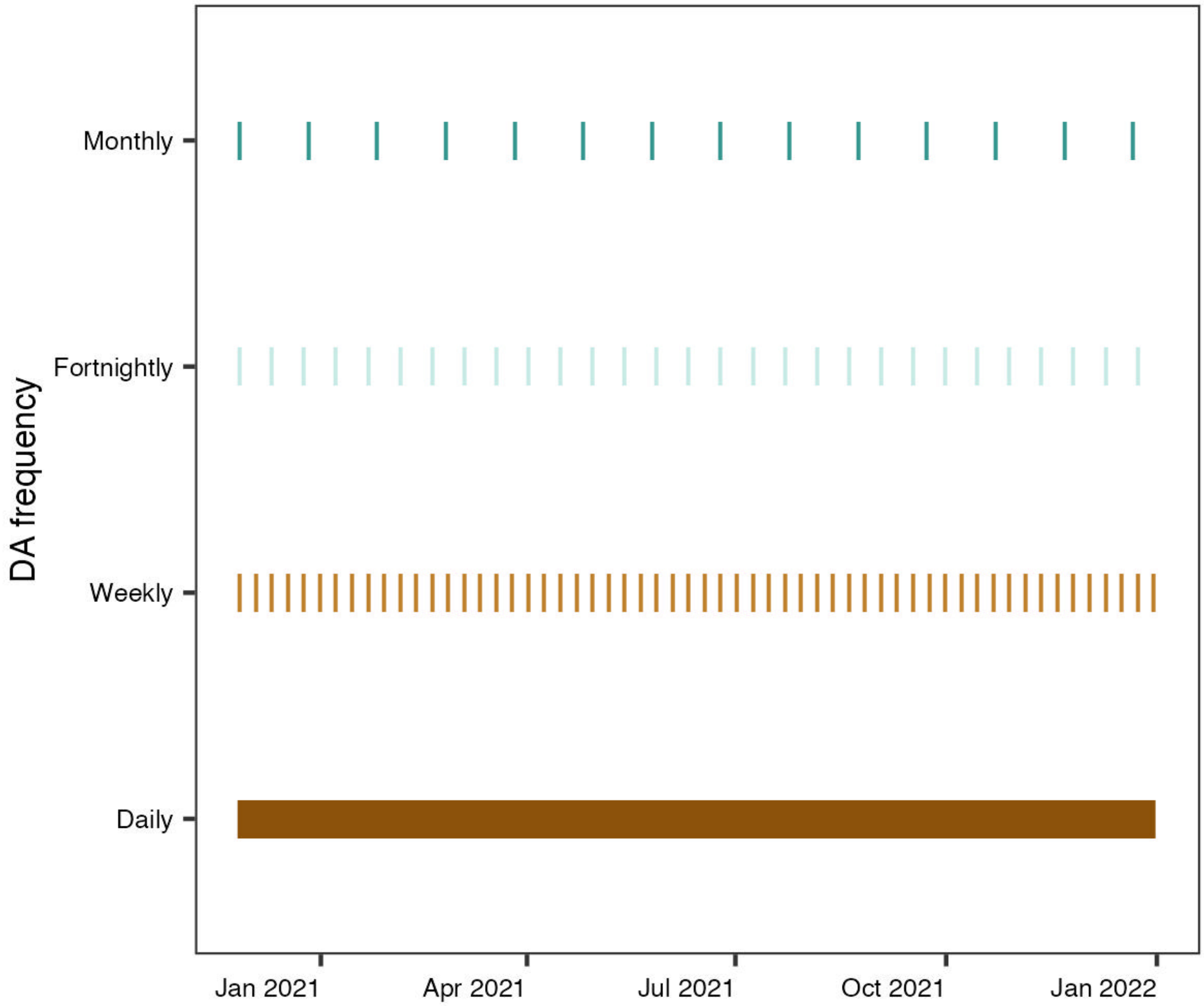


Figure S1: Frequencies for daily, weekly, fortnightly, and monthly data assimilation (DA); lines indicate the dates when DA occurred. For example, daily DA occurred every day from 27 November 2020 to 31 December 2021, whereas monthly DA occurred 14 times during the 14-month period of November 2020 to January 2022.

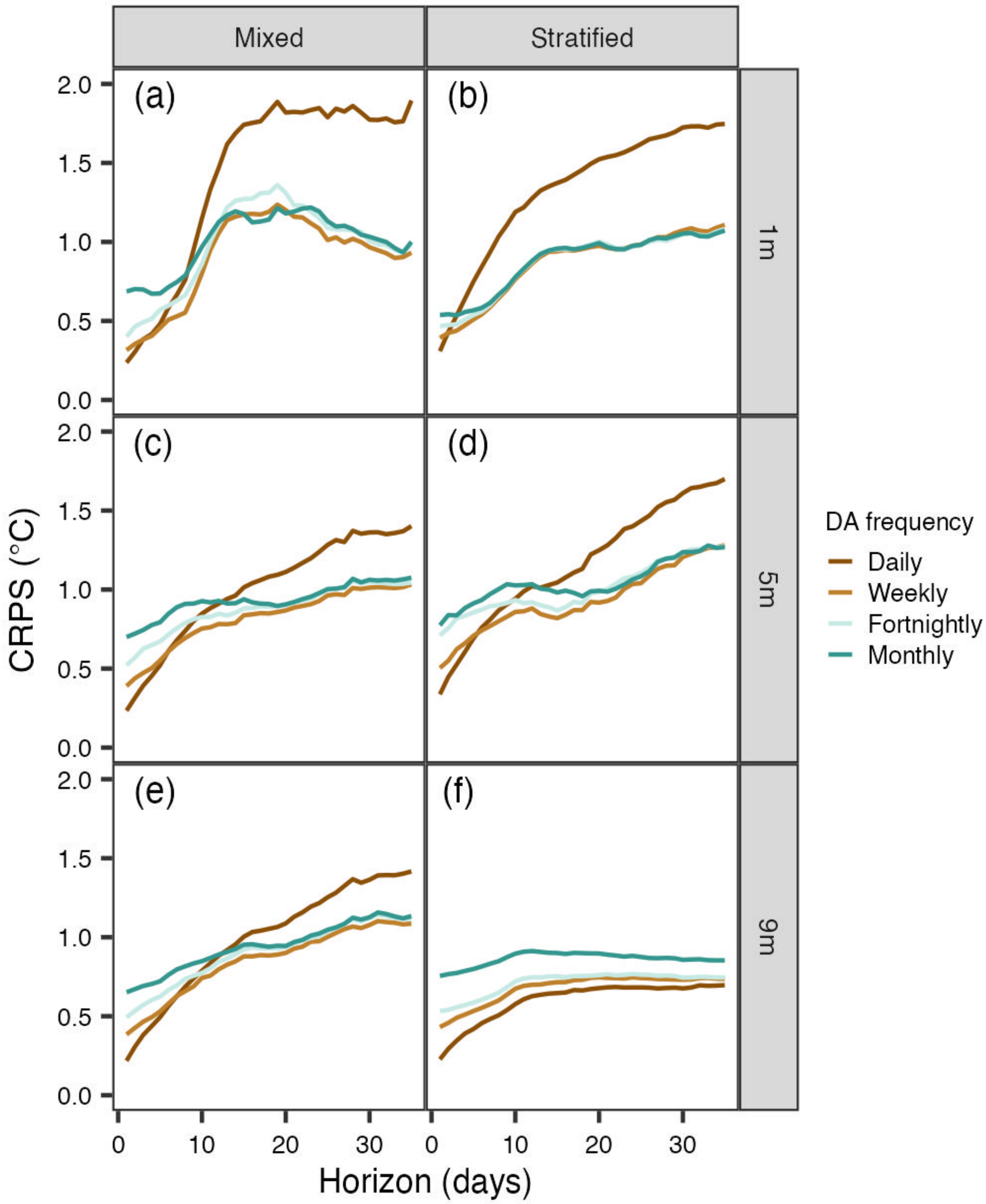


Figure S2: Water temperature forecast continuous ranked probability score (CRPS) across different depths (1 m: a, b; 5 m: c, d; and 9 m: e, f) and horizons in Beaverdam Reservoir during the mixed (a, c, e) vs. stratified (b, d, f) periods. Each grouping of bars represents a different data assimilation (DA) frequency (daily, weekly, fortnightly, monthly). Depths are indicated in the right y-axis labels; horizons are depicted by colored boxplots.

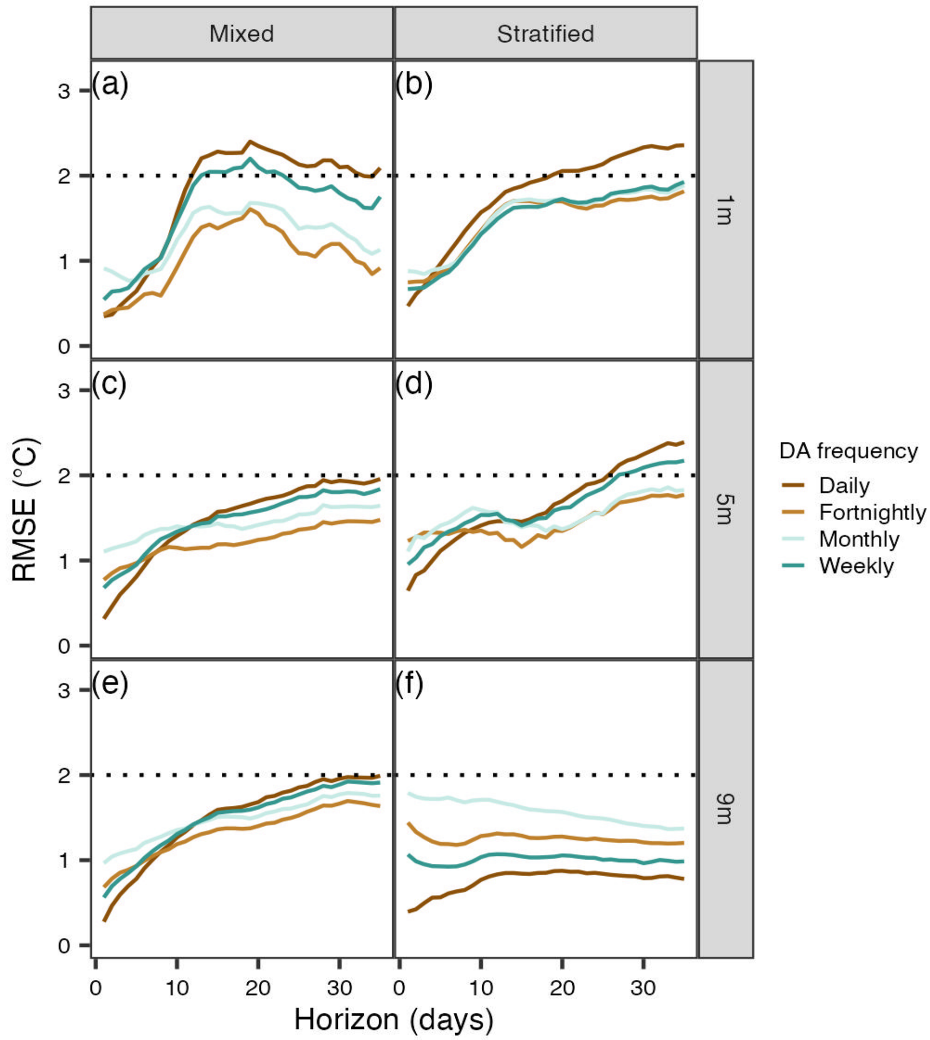


Figure S3: Root mean square error (RMSE) calculated from comparing water temperature observations with forecasts that did not include initial conditions uncertainty for 1-35-day-ahead forecasts in Beaverdam Reservoir during the mixed (a, c, e) vs. stratified (b, d, f) periods at 1 m (a, b), 5 m (c, d), and 9 m (e, f). RMSE for each forecast horizon was averaged across the 365-day forecast period (1 January - 31 December 2021). Colored lines correspond to different data assimilation (DA) frequencies (daily, weekly, fortnightly, and monthly); dotted horizontal lines depict the 2℃ RMSE threshold for skillful forecasts.

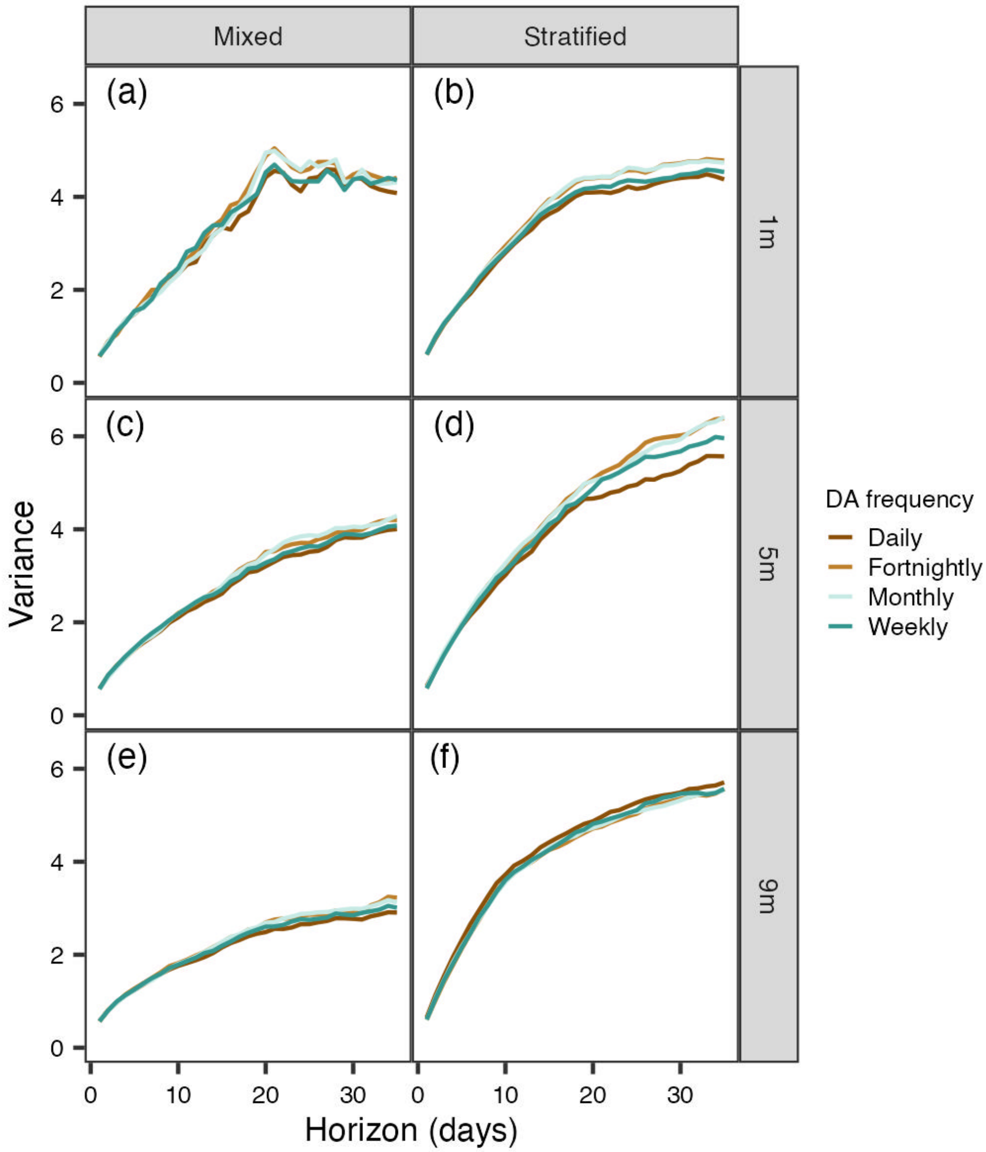


Figure S4: Variance for forecasts that did not include initial conditions uncertainty across horizons (1-35 days ahead) in Beaverdam Reservoir during the mixed (a, c, e) vs. stratified (b, d, f) periods for 1 m (a, b), 5 m (c, d), and 9 m (e, f). Variance for each forecast horizon was averaged from all 365 forecasts generated during the forecast period (1 January - 31 December 2021). Colored lines correspond to different data assimilation (DA) frequencies (daily, weekly, fortnightly, and monthly).